AFRL-RV-HA-TR-2007-1191

Seismic Source Locations and Parameters for Sparse Networks by Matching Observed Seismograms to Semi-Empirical Synthetic Seismograms
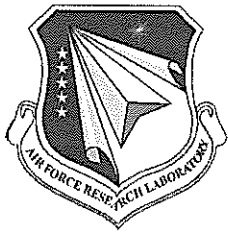
David H. Salzberg
Margaret E. Marshall

Science Applications International Corporation
1710 Saic Drive
McLean, VA 22102

Final Report

6 December 2007

AIR FORCE RESEARCH LABORATORY
Space Vehicles Directorate
29 Randolph Road
AIR FORCE MATERIEL COMMAND
Hanscom AFB, MA 01731-3010

NOTICE AND SIGNATURE PAGE

AFRL-RV-HA-TR-2007-1191 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

//Signed//   //Signed//

ROBERT RAISTRICK   PAUL TRACY, Acting Chief
Contract Manager   Battlespace Surveillance Innovation Center

REPORT DOCUMENTATION PAGE [example]

1. REPORT DATE *(DD-MM-YYYY)* 11-30-2007	2. REPORT TYPE Scientific Report - Final	3. DATES COVERED *(From - To)* 3-31-2005 to 10-31-2007

4. TITLE AND SUBTITLE Seismic Source Locations and Parameters for Sparse Networks by Matching Observed Seismograms to Semi-Empirical Synthetic Seismograms	
	5a. CONTRACT NUMBER FA8718-05-C-0019
	5b. GRANT NUMBER N/A
	5c. PROGRAM ELEMENT NUMBER 62601F
6. AUTHOR(S) D. H. Salzberg M. E. Marshall	5d. PROJECT NUMBER 1010
	5e. TASK NUMBER SM
	5f. WORK UNIT NUMBER A1

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Science Applications International Corporation 1710 SAIC Dr McLean, VA 22102	8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Research Laboratory 29 Randolph Rd. Hanscom AFB, MA 01731-3010	10. SPONSOR/MONITOR'S ACRONYM(S) AFRL/RVBYE
	11. SPONSOR/MONITOR'S REPORT NUMBER(S) AFRL-RV-HA-TR-2007-1191

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

In this study, we have used a semi-empirical approach to computing synthetic seismograms to demonstrate full-waveform earthquake location and provide size estimates of one nuclear explosion (2006 North Korea Event). The overall approach is to use a known (reference) event to empirically characterize the propagation by comparing the reference event to a synthetic seismogram computed for appropriate. This empirical correction is then applied to a synthetic seismogram computed for a different location and/or source mechanism. Three study areas were used for locations: Central California, Nevada Test Site, and Lop Nor. Events were located to within 5 km of ground truth locations (though in semi-major axes of the error ellipses were greater than 5 km). The approach was also used to estimate the yield of the Oct. 9, 2006 North Korea Nuclear Test using a small 1.2 - 1.5 ton explosion as a reference event. The resulting yield was a chemical equivalent of 372-465 tons.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON Robert Raistrick
a. REPORT UNCLASSIFIED	b. ABSTRACT UNCLASSIFIED	c. THIS PAGE UNCLASSIFIED	SAR	35	19b. TELEPHONE NUMBER *(include area code)* 781-377-3726

Table of Contents

Figures

Tables

1. SUMMARY

In this study, we have used a semi-empirical approach to computing synthetic seismograms. These synthetics were used to demonstrate full-waveform earthquake location and provide size estimates of one nuclear explosion (2006 North Korea Event). The overall approach is to use a known (reference) event to empirically characterize the propagation by comparing the reference event to a synthetic seismogram computed for the appropriate path and mechanism. This empirical correction can then be applied to a synthetic seismogram computed for a different location and/or source mechanism. This is our "semi-empirical synthetic seismogram." Two methods were used to empirically characterize the propagation: a broad-band approach, which provides very precise locations when the new event and reference events are close (within 10 km) and also provides the event size estimates. A second, long-period range dependent solution was used to provide reasonable locations when the events were further apart.

Three study areas were used to validate the location approach: Central California, Nevada Test Site, and Lop Nor. In all cases, events were located to within 5 km of ground truth locations (though in some cases, the semi-major axes of the error ellipses were greater than 5 km) . This accuracy was obtained even when the event separations were greater than 50 km, and when the source mechanisms differed between the reference and new events when the range dependent approach was used. The broad-band was used for a few Central California Events; in those cases, the events were located to within one sample space, which corresponds to ½ km if the two events were within a few kilometers of each other.

In comparison, the North Korea Nuclear explosion of October 9, 2006 occurred in a seismically isolated region (no nearby earthquakes). However, an active source experiment for wide-angle reflection and refraction near the border in China (Song et al, 2007) provides some sort of calibration. While the yield China experiment was small (1.2 - 1.5 tons), a good signal was recorded at the station MDJ for one particular explosion. This event, roughly ½ way between the North Korean Nuclear Explosion and MDJ, provides our reference event for our yield estimate of the North Korean Nuclear Explosion, which is a chemical equivalent of 450 tons.

2. INTRODUCTION

In this study, we examine the whole-waveform processing techniques using semi-empirical synthetic seismograms. The overall approach is to use a known (reference) event to empirically characterize the propagation by comparing the reference event to a synthetic seismogram computed for appropriate. This empirical correction can then be applied to a synthetic seismogram computed for a different location and/or source mechanism. This is our "semi-empirical synthetic seismogram." The goal of study presented here is to demonstrate the usefulness of the semi-empirical synthetic

waveforms with earthquake location as well as event size estimates. As such, we require high quality "Ground Truth" locations for earthquakes recorded at two or more regional stations.

For event location, three study areas were used: Central California, Nevada Test Site, and Lop Nor.   The ground truth for the three data sets are the Berkeley Seismic Network locations (`http://seismo.berkeley.edu/`), the Nuclear Explosion Database (Yang et al, 2000), Fisk's (2002) study of the Lop Nor region, and from the International Seismological Centre (2001).   The quality of the solutions from the first three sources are GT1 or better; the one event from the ISC was presumably GT5. The reference events for the three regions are a Parkfield, Ca aftershock, the BEXAR explosion, and a Lop Nor Explosion.

For the event size example, we applied it to one event: the North Korean Nuclear Explosion of October 9. 2006. The reference event used was a 1.2- 1.5-ton explosion that was part of a wide-angle refraction/reflection experiment (Song et al, 2007). To find the relative amplitude (or chemical equivalent yield), the energy envelopes of the two events are compared and scaled.

3. TECHNICAL APPROACH

Matching observed waveforms with synthetic waveforms has been used for some time in ocean acoustics to provide robust estimates of underwater source locations in range, bearing and depth (Baggeroer et al., 1993) using synthetic pressure fields.   In such matched field processing, the pressure time series recorded at each element of a tightly spaced an array of hydrophones correlated with the synthetic pressure field. The variation is the wave field resulting from interference patterns may be unique to the source location. By searching over a grid of potential sources, the optimal source location is identified by the maximum in correlation. This approach is ideally suited for the ocean environment since the medium properties are well constrained.

In contrast, the properties of the solid earth are not known at the detail required to perform matched field processing on seismic arrays for source location on a global scale. However, matching synthetic waveforms has been effectively applied in seismology to techniques such as inversion of the source and inversion of velocity models (Burdick and Langston, 1977).   Waveform matching for source properties has been limited to regions with well constrained velocity models such as teleseismic body-waves which travel mostly in the lower mantle (Langston, 1981), long period surface waves (Romanowicz, 1981), and normal modes (Dziewonski et al., 1981). In these three examples, most of the energy propagates within the lower mantle. However, matched waveform processing for source locations, has only been successfully applied when the overall velocity is well constrained (Pulliam et al., 2000).

SAIC proposed to overcome the uncertainty in the velocity models by generating semi-empirical synthetic seismograms. Empirical and semi-empirical approaches for modeling

waveforms have been used to model waveforms from large events using smaller events as Green's functions (Wu, 1978). This approach, however, breaks down if the two events are neither co-located nor have similar mechanisms. Others (e.g., Summerville et al., 2000) use semi-empirical synthetic seismograms, i.e., synthetic seismograms convolved with empirically determined source time functions, to estimate the ground motion of hypothetical events to assess earthquake hazards. Salzberg (1996) developed a semi-empirical technique for fundamental mode surface waves which allowed synthetic seismograms to be computed when the reference event is at a different source location and has a different mechanism. This technique provided coherent surface-waves at periods as short as 15 seconds, and allowed for the determination of moment tensor source mechanisms for small events (M > 4.7).

SAIC's approach is to find the semi-empirical synthetic seismogram computed across a dense grid of possible locations that best matches the observed seismic waveforms. For each grid location, two or more 3-component waveforms are inverted for the moment tensor using the semi-empirical synthetic waveforms. The resulting minimum in the residual (χ^2 norm) from the moment tensor inversion results at the grid points yields the optimal location. Figure 3 shows a conceptual two-dimensional illustration of the proposed methodology, although the actual grid will be evenly spaced in three dimensions providing resolution of the latitude, longitude and depth.

3.1. Semi-Empirical Synthetics

In the synthetic test shown in Figures 5 and 6, synthetic waveforms were compared with other synthetic waveforms, all computed with the same velocity model. When comparing the synthetic waveform with real data, a match is not possible unless the velocity model is well defined. Furthermore, errors in the synthetic waveforms translate into errors in the resolved location and source parameters. Consequently, the primary limitation in matched waveform processing is the ability to produce high-fidelity simulations based on an idealized environmental model. One approach that can be used to improve the quality of the synthetic waveforms is to empirically determine the propagation portion of the synthetic waveform, yet still use the theoretical excitation (Salzberg, 1996; Velasco et al., 1994). These semi-empirical Green's Functions can be used to characterize the seismic wavefield recorded from sources in a reasonably homogeneous source region at a seismic station or array.

Semi-empirical Green's Functions are unnecessary if the source-to-receiver velocity structure is known perfectly, as full waveform synthetic seismograms are sufficient for matched waveform processing. However, only high-quality approximations of the velocity model can be determined. As such, any synthetic waveforms generated will only approximate the observed waveform written as:

*observed (t) = u(t) * Δu(t)* **(1)**

where * represents the convolution operation, u is the synthetic waveform, and *Δu* is the mismatch of the data and synthetic represented as a filter. *Δu* corresponds to a systematic

bias caused by model mismatch and an incoherent portion caused by random noise. The systematic portion of Δu can be written as:

$$\Delta u\,(\omega) = \Delta p(\omega)\cdot \Delta x(\omega) \tag{2}$$

where Δp is the mismatch caused by inaccuracies in the propagation, and Δx is the mismatch in the source excitation. The synthetic waveform, u, can be written as (Mendiguren, 1977):

$$u(r,\theta,h,\omega) = P(r,\omega)\cdot s(\omega)\cdot \sum_{i=1}^{5} H_i(\omega,\theta,h)\cdot m_i \tag{3}$$

where u is the far-field displacement, s is the source function, P is the propagation from source to receiver, H_i is the excitation function corresponding to m_i, and m_i is the ith element of the moment tensor. Transferring Equation (1) into the frequency domain, and combining with Equations (2) and (3) can be combined to represent the observed seismogram as:

$$observed(\omega) = P(r,\omega)\cdot \Delta p(\omega)\cdot \left[s(\omega)\cdot \sum_{i=1}^{5} H_i(\omega,\theta,h)\cdot m \right]\cdot \Delta x(\omega) \tag{4}$$

In this formulation, the propagation term is mathematically separable from the source terms which include the moment tensor, excitation, and time functions. After deconvolving the source terms from the reference waveforms of known source mechanism, location, and depth, the source-to-receiver propagation is isolated as:

$$P(r,\omega) = \frac{reference(\omega)}{\left[s(\omega)\cdot \sum_{i=1}^{5} H_i(\omega,\theta,h)\cdot m_i \right]\cdot \Delta x(\omega)\cdot \Delta p(\omega)} \tag{5}$$

If a second event occurs close to the first event, then it reasonable to assume that the propagation from the second event to the receiver will be similar to the first event. The reference waveform can then be transferred to the mechanism and depth of the new waveform by substituting the empirical propagation determined in Equation (5) into observed waveform formulation in Equation (4). This result gives new waveform as

$$new(\omega) = \frac{reference(\omega)}{\left[\tilde{s}(\omega)\cdot \sum_{i=1}^{5} \tilde{H}_i(\omega,\theta,\tilde{h})\,\tilde{m}_i \right]\cdot \Delta\tilde{x}(\omega)\cdot \Delta\tilde{p}(\omega)}\cdot \left[\sum_{i=1}^{5} H_i(\omega,\theta,h)\cdot m_i \right]\cdot \Delta x(\omega)\cdot \Delta p(\omega) \tag{6}$$

where ~ over a variable indicates the reference waveform.

Accurate knowledge of the source moment tensor and excitation terms implies that the Δx terms are approximately 1. Furthermore, since the two events are in close proximity and assumed to have the same propagation, or $\Delta \tilde{p} = \Delta p$. Thus the new semi-empirical synthetic waveform is:

$$new(\omega) = \frac{reference(\omega)}{\left[\tilde{s}(\omega) \cdot \sum_{i=1}^{5} \tilde{H}_i(\omega,\theta,\tilde{h}) \tilde{m}_i\right]} \cdot \left[\sum_{i=1}^{5} H_i(\omega,\theta,h) \cdot m_i\right] \tag{7a}$$

$$new(\omega) = \frac{reference(\omega)}{synthetic_{reference}(\omega)} \cdot synthetic_{new}(\omega) \tag{7b}$$

Conceptually, the formulation in (7b) assumes that the mismatch between the reference waveform and its synthetic waveform is identical to the mismatch between the new waveform and its corresponding synthetic waveform.

Proposed wave number (K-space parameterization)
In this section, the *k-space* parameterization will be described. At the end of the derivation, the resulting formulation is equation (2).

Parameterizing our classic formulation, shown in equation (4) in terms of amplitude and phase yields:

$$\Delta u(\omega, r_o) = \left|\frac{s(\omega)}{u(\omega, r_o)}\right| \cdot e^{i\left(\phi_s - \phi_{u_o}\right)} = \Delta A(\omega) \cdot e^{i\Delta\phi} \tag{8}$$

This approach works well if the events are close together, or if there is a dominant phase in the data (e.g., Rayleigh). In the case where the spatial separation of the events approaches or is greater than the wavelength of the highest frequency energy used, the increasing time separation of the body waves and phase mismatch of the surface waves results in increased misfit, as increases occur in the separation of the reference and the second event.

An alternative, range dependant phase parameterization will provide better agreement of the waveforms in these cases. Treating the waveform as a sum of propagating modes, the modes can be represented as a range dependent amplitude and range dependent phase, or

$$u(\omega, t, k, r) = \sum_{n} A_n(\omega, r) \cdot e^{i(k_n r - \omega t)} \tag{9}$$

where, u is the displacement, A is the amplitude, k is the wavenumber, and n is the mode.

Using this formulation for both the data and the synthetic waveform (and assuming that the modes of the data can be separated), equations (8) and (9) can be combined to parameterize the phase misfit to terms of the wavenumber and distance, or

$$\Delta\phi = \Delta\phi_{reference} - \Delta\phi_{synthetic} = \left(k_{reference} \cdot r_{reference} - \omega \cdot t_{reference}\right) - \left(k_{synthetic} \cdot r_{synthetic} - \omega \cdot t_{synthetic}\right)$$
(10)

Assuming the reference time for the data and the synthetic is the same, and the synthetic was computed for the observed distance, then

$$\Delta\phi = \left(k_{reference} \cdot r - \omega \cdot t\right) - \left(k_{synethic} \cdot r - \omega \cdot t\right) = \left(k_{reference} - k_{synethic}\right) \cdot r = \Delta k \cdot r$$
(11)

For each propagating mode, assigning the reference event range to be r_o, allows the empirically filtered waveform for the specific propagating mode at range r to be expressed as:

$$u_e\left(\omega, r\right) = \frac{s\left(\omega\right)}{u\left(\omega, r_o\right)} \cdot e^{i\Delta k\left(r - r_o\right)} \cdot u\left(\omega, r\right)$$
(12)

With real data, it is not feasible to separate each propagating mode exactly. However, it is possible to apply time-frequency windowing functions to isolate the body waves from the surface waves.

4. RESULTS AND DISCUSSION

The primary focus of this study (and the method) was to provide accurate locations using sparse networks based on our semi-empirical methodology. A secondary study was performed to asses the size (yield) of the Oct 2006 North Korean Nuclear Test. As such, we will first present the location processing results, followed by a discussion of the approach and results for the yield estimate.

4.1. Location

Our approach to location is to the semi-empirical synthetic seismograms across a grid, then compare those waveforms to the observed waveform (in a χ^2 sense), and to find the grid point with the best fit (minimum in χ^2 space). That point is considered the optimal source location. The variability of χ^2 is used to estimate the uncertainty; we use 2x the minimum of our χ^2 estimate to assign the uncertainty estimates, as characterized by an error ellipse. Key to the approach is the generation of the synthetic seismogram. For this study, the wavenumber integration approach implemented by Herrmann (2002) is used. With this approach, a synthetic can be computed if the mechanism and velocity models are known. The mechanisms are obtained through source inversion , the Berkeley Moment Tensor Catalog (http://seismo.berkeley.edu/~mike/solutions.new), and the Global Centroid Moment Tensor catalog (http://www.globalcmt.org/). The models are obtained using a Monti-Carlo Simulation as described below.

4.1.1. Development of an Optimal Velocity Model

Even though high-quality velocity models are available for both Central California and the Nevada Test Site, since we are developing an approach that should be portable, a robust approach was used to develop path specific velocity models. Our approach was to execute Monti-Carlo simulations by randomly varying material properties: Vp, Vs, and the density, as shown in Figure 1. Our approach was to compute 3000 models, then compute long-period synthetics for our reference event. The optimal model used for Central California is shown as red in Figure 1, is the model with the minimal misfit; the RMS vs. model number is shown in Figure 2. A comparison of the observed and synthetic waveforms is shown in Figure 3. Note that this model worked best for all of the stations.
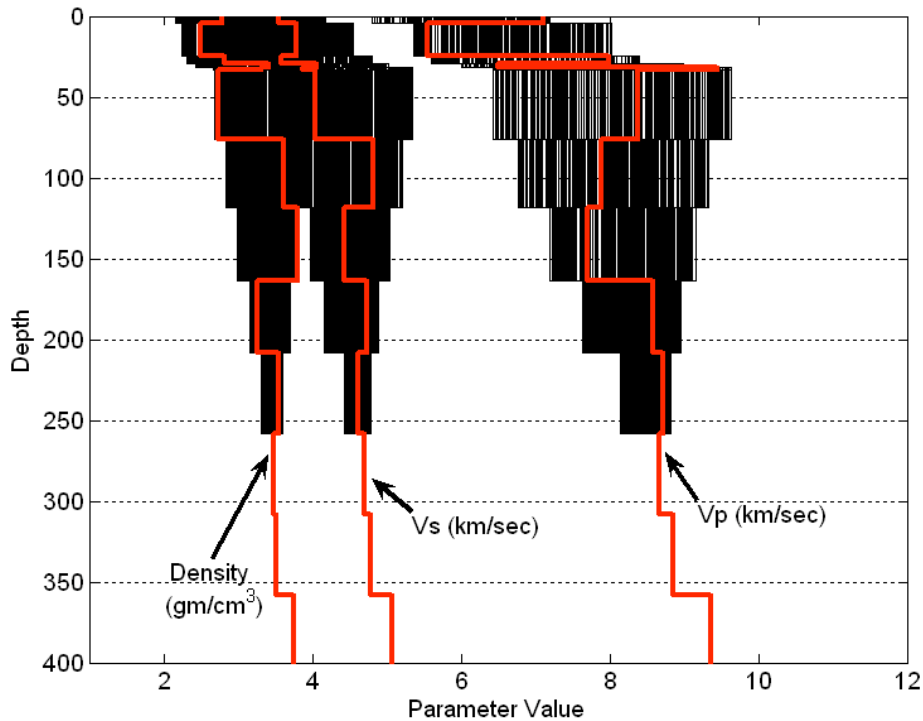
Figure 1: The optimal velocity model (red) compared with the range of velocity models generated for the Monte Carlo simulation (Black). The models were generated by randomly varying the material properties for shallow layers, while constraining the deeper models to IASPEI.
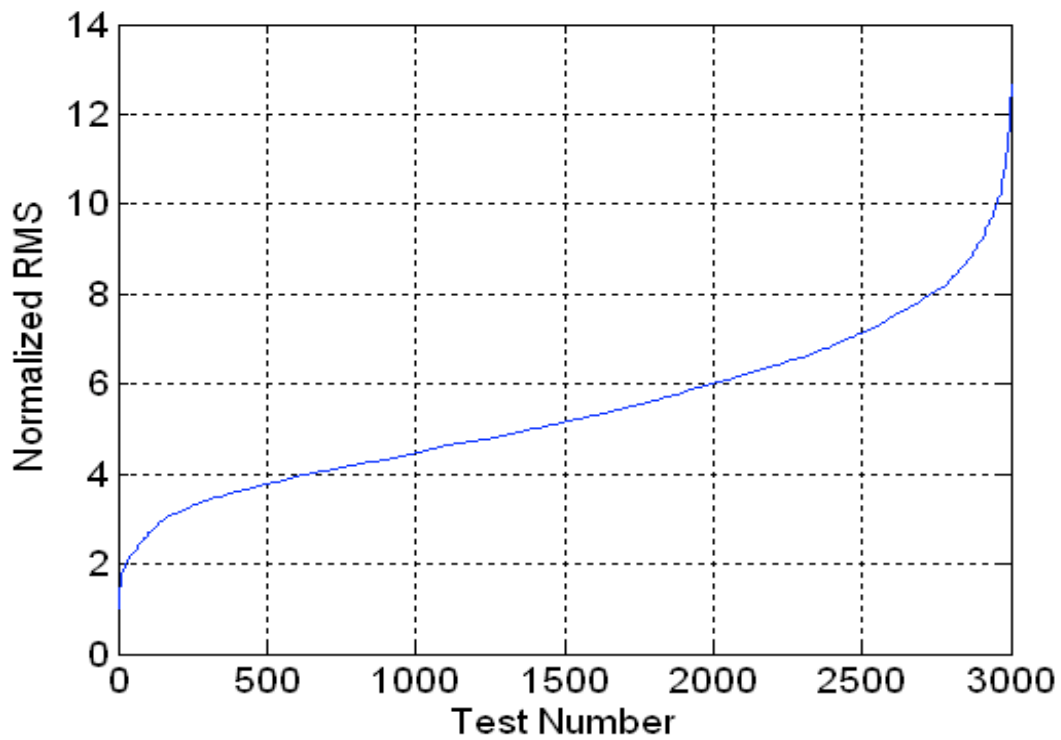
Figure 2: The normalized root mean square difference between the observed long-period waveform at ELK and the synthetic waveform for each model. Note that the optimal model is 13x less than the worst model, and about 5x better than the median model.
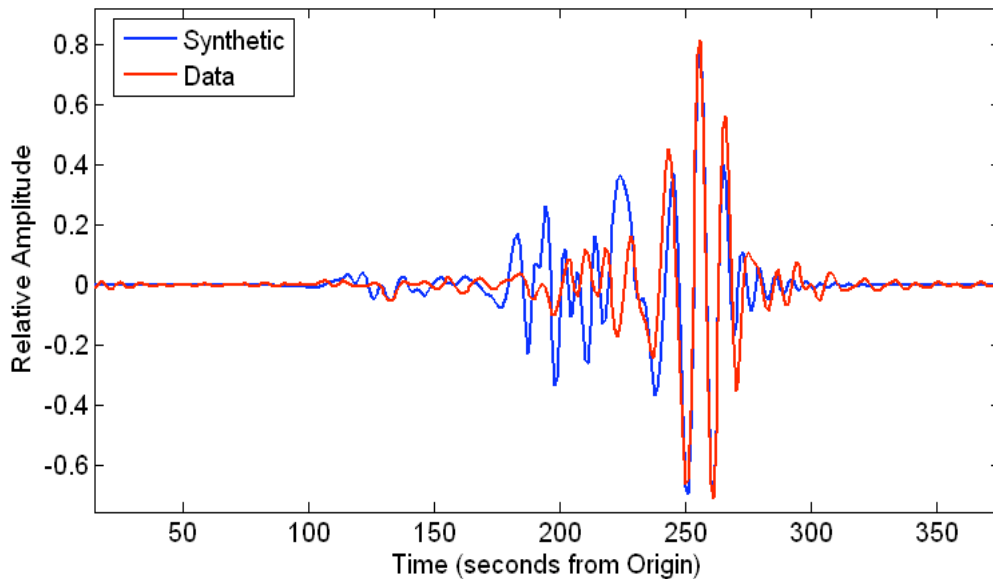
Figure 3: The band-passed (.01-.2 Hz) synthetic waveform for the optimal velocity model compared with the synthetic waveform shows excellent agreement between the synthetic and observed Rayleigh waves. In addition, the arrival times of the body phases also agree. The mismatch in amplitude and shape may result in the near nodal nature of the waveform at ELK.

8

4.1.2. Broad-Band Test Case: Parkfield, California

When we first conceived of the semi-empirical technique for event location, the concept was a broad-band, full-waveform approach. Our initial example was well suited for this: our reference event was an aftershock of the 2004 Parkfield Earthquake (2004/09/29 17:10:04 GMT), and the new event (2004/09/30 18:54:29 GMT) was a nearly co-located event. The event was located first in range and depth for the three stations (ELK, PAS and CMB) over a ½ km by ½ km grid spacing in range and depth. The mechanisms for the two events were from the Berkeley Moment Tensor Catalog. The residual vs. range and depth along the waveform comparison at the optimal depth are shown in Figure 4. From this figure, it is apparent that the optimal depth and range closely agree with the optimal depth from the grid search. Furthermore, the waveforms show remarkable agreement. In addition, when the three stations are combined to determine the event hypocenter, the optimal solution, as shown in residual space, shows remarkable agreement with the ground truth location (Figure 5), and is much better than the CTBTO IDC REB location.
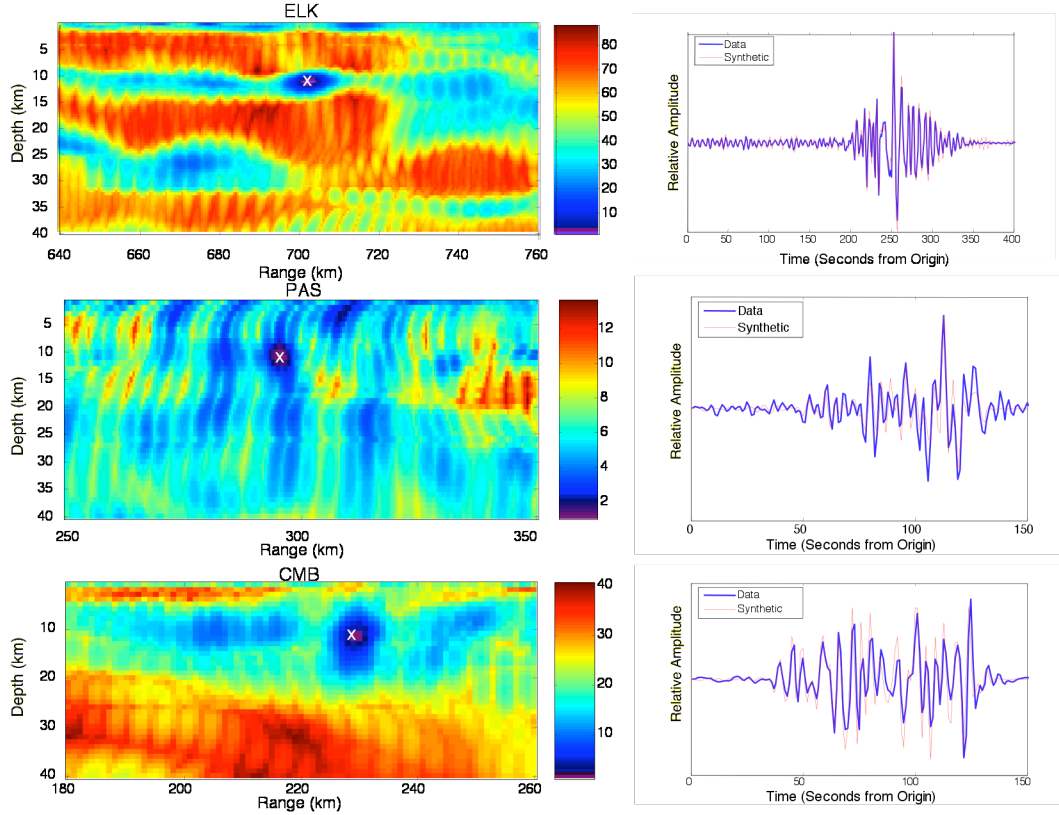
Figure 4: For each station, the RMS surface over depth and range over the frequency band of 0.25 to 0.01 Hz is shown on the left, and the comparison between the semi-empirical synthetic and the observed is shown on the right.
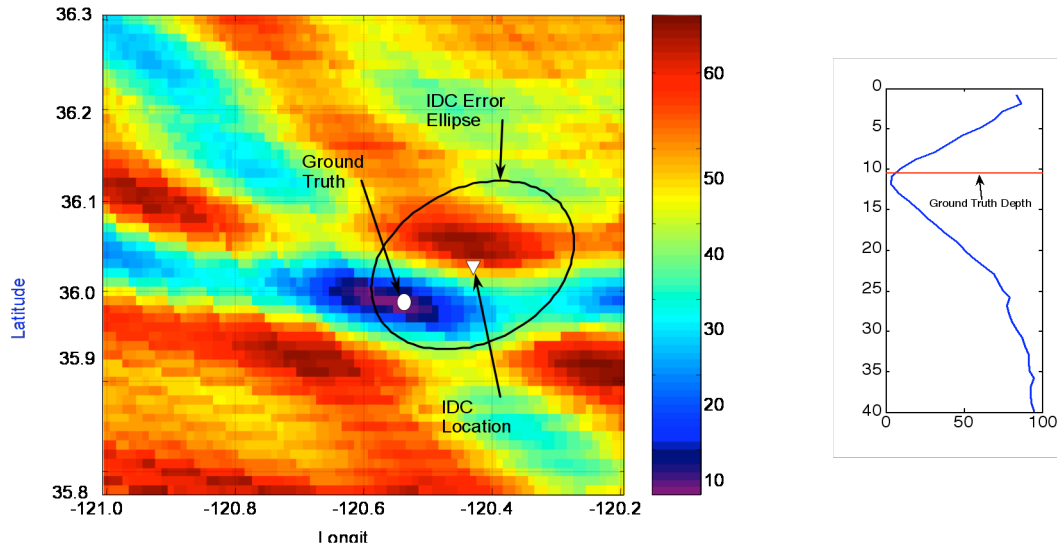
Figure 5: The event location by combining the results for the three stations shows sharp minima (left), and the minimum in RMS is within one or two pixels of ground truth. Our solution is much closer to the GT location than the IDC's location. Furthermore, RMS vs. Depth (lower right) shows that the depth is well resolved.

In order to examine how the method works with smaller events with greater event separation, we applied the technique to an Mw= 4.1 that that was probably triggered by the Parkfield Earthquake. Of the three stations, only CMB had an observable long-period (f <1 Hz) signal. Figure 6 shows the RMS over range and depth, and the corresponding synthetic waveforms for a variety of frequency bands. The full-band RMS over depth and range showed no minimum in RMS (high-frequency cutoff of 0.45 Hz, lower-most image). The corresponding waveform shows poor agreement with the observed data. Moreover, only the lower frequency energy (f< 0.1 Hz; top figure) shows any agreement between data and the semi-empirical synthetic waveform. However, at the lower frequencies, we are able to resolve the depth and location (though not that well), though with an uncertainty and error unacceptably large.

Additional investigation regarding the failure of the method for this (smaller) event indicates that the difficulty arises from 6 second microseismic noise. In fact, comparing the broadband spectrum of the signal to post signal background noise in Figure 7 (left) clearly shows that there is significant SNR at frequencies below 0.1 HZ and above 0.5 Hz. From the time-series, shown on the right side of Figure 7, it is clear that, while the long-period body-waves are in the noise, they are clearly visible in the broadband waveforms. As such, an approach that utilizes the broad-band (and high frequency) body waves along with the lower frequency surface waves would be optimal. However, one must be careful, because as the processing frequency increases, then source duration and finiteness would impact the semi-empirical synthetic seismograms.
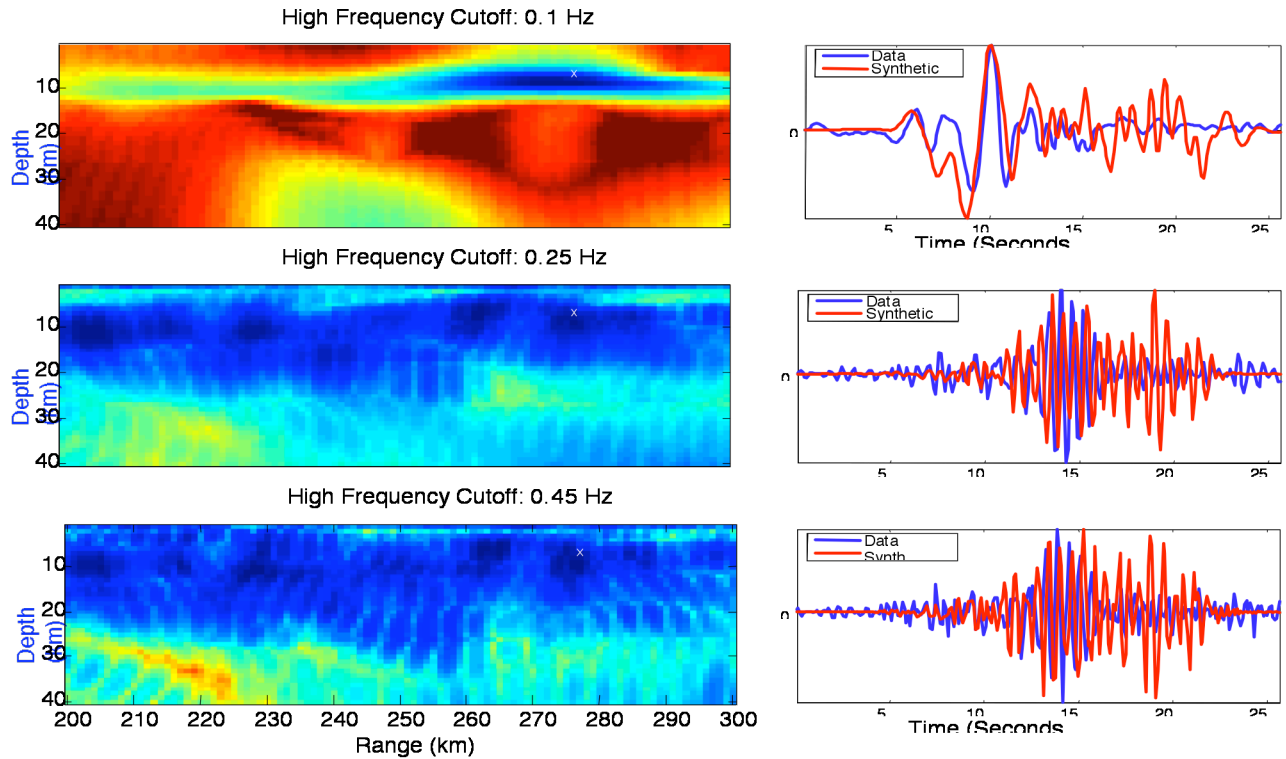
Figure 6: The residual over depth and range for station CMB for an event that occurred on October 2, 2004, 12:22:09. The event was about 50 km SW of the reference event, shown on he left. On the right are waveforms at the 'best'. From this, it is apparent that a solution in range and depth was not obtained using higher frequency data (Bottom two figure). However, low-frequency data (top figure) shows a broad minimum.
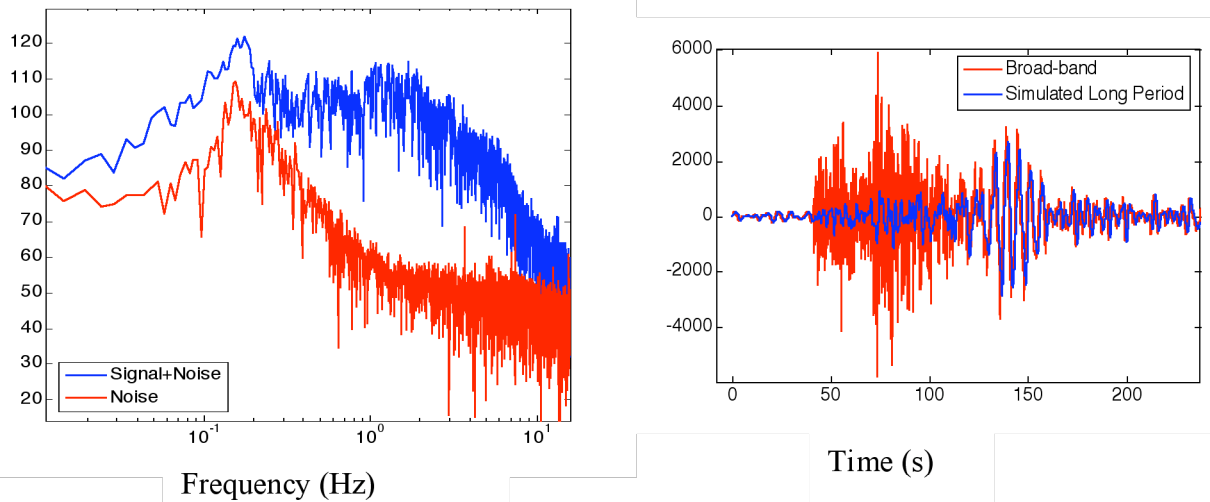
Figure 7: The broad band frequency response (left) and a comparison of the broad band time series with the long-period time series (right). Note that at low and high frequencies, there is significant signal-to-noise. However, in the miscroseismic band (.1 to .5 Hz), the signal and noise are approximately equal.

4.1.3. Range Dependent Examples

As previously discussed, a primary issue when applying the broad-band technique is longer period signal-to-noise as well as the fact that a single semi-empirical term is computed for all phases, even though the dispersed seismic waveform strongly suggests that a phase correction for the P-wave (for example) would not be appropriate for other phases (S, Surface waves). Using a range-dependent parameterization of the semi-empirical synthetic seismograms solves this problem. In this approach, we isolate specific phases (or modes), compute the empirical correction term for each phase (or mode). As a practical matter, this is limited to the longer period energy (f<1 Hz) because of potential contamination from source duration. Thus, the only phase that is usable are usable are the surface waves. As such, the approach used is to isolate and compute semi-empirical synthetic seismograms for the long period (f< 0.12) Hz Rayleigh waves. The high frequency energy in the P-waves (e.g.,Figure 7) is used to provide a timing reference. At each range and depth grid point, the semi-empirical synthetic is shifted such that its predicted P-wave arrival time matches the observed P-wave arrival times.

The approach has been applied to three regions: The Nevada Test Site (NTS), Central California, and Lop Nor, China.

4.1.3.1. NTS and surrounding areas

The range-dependent approach is first applied to region around the Nevada Test Site. This location is well suited to validate the approach, as there are both earthquakes and explosions over a 60 km wide area. Furthermore, the locations of the explosions are known to GT0 (Yang et al, 2000). As such, four explosions and one earthquake (Little Skull Mountain) were processed using data from the seismic stations, ANMO and PAS, as shown in Figure 8, and listed in Table 1.

The per-event processing results, shown in Figures 9-13, show that the Rayleigh wave, when aligned observed P-waves are aligned to the predicted P-wave times, allows us to locate events to around GT5 precision, even when the mechanism differs between the reference event and the new event (e.g., Little Skull Mountain Earthquake, Figure 8, Table 1). When comparing the result with the GT, as shown in Figure 14, it is clear that all of the events are located accurately (~GT5), even when the event separate is large (60 km). In fact, looking at the errors compared with separation from the reference events (**Figure 15**) clearly shows that, in all cases, our best location is within 4 km of ground truth, though the location uncertainty as measured by the error ellipse was on the order of 7 km for the two distant events.
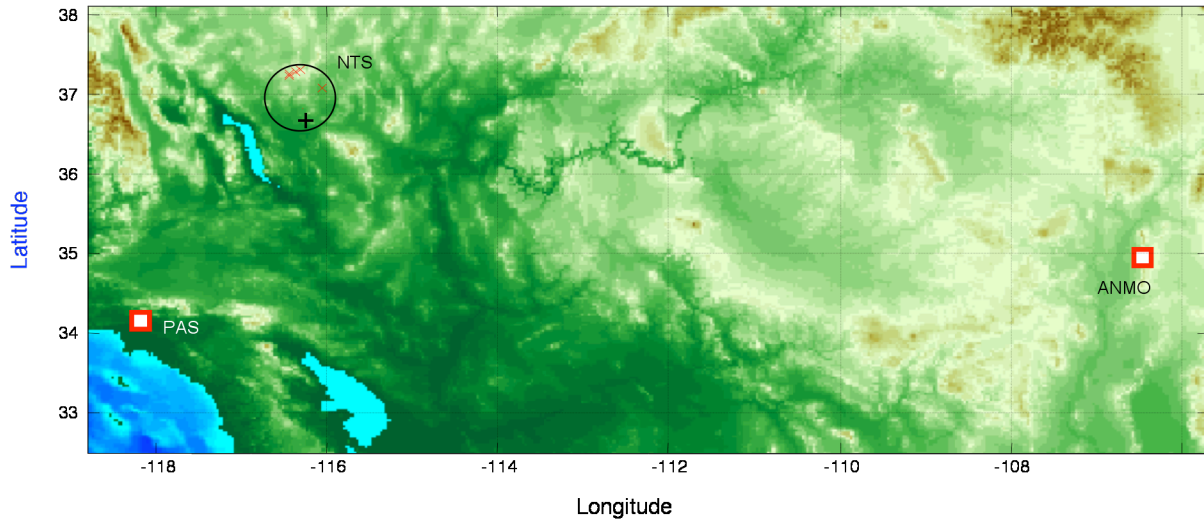
Figure 8: The geographic for the NTS dataset.. Note that smaller explosions were not used because of poor signal to noise at more distant station ANMO.

Table 1: The events used in the study. The first 5 listed are nuclear explosions, the last event is a normal faulting earthquake. The comment field gives the explosion name, and identifies the reference event.

Date	Lat	Lon	Depth (km)	m_b	Comment
1991/04/04 19:00:00	37.2961	-116.3129	0	5.6	BEXAR
1991/04/16 15:30:00	37.2454	-116.4416	0	5.4	MONTELLO: Reference
1991/09/14 19:00:00	37.2256	-116.4281	0	5.5	HOYA
1991/10/18 19:12:00	37.0634	-116.0453	0	5.2	LUBBOCK
1992/03/26 16:30:00	37.2724	-116.3598	0	5.5	JUNCTION
1992/06/29 10:14:22	36.7140	-116.3068	9.2	5.6	Little Skull Mountain Earthquake

14
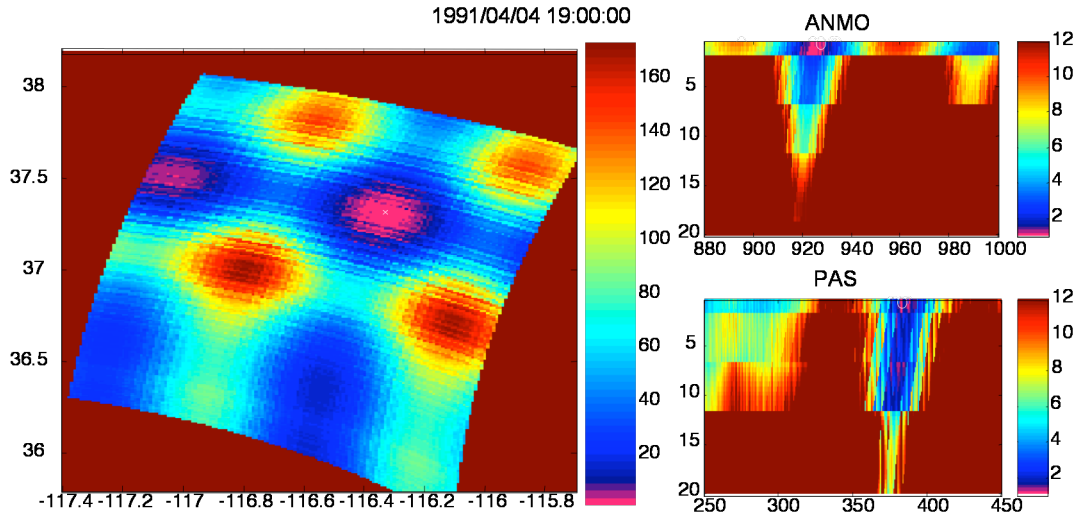
Figure 9: Range vs. Depth for ANMO and PAS (right) and a map view of the residual at the optimal depth (left) for the BEXAR explosion. The Ground truth location is the white X.
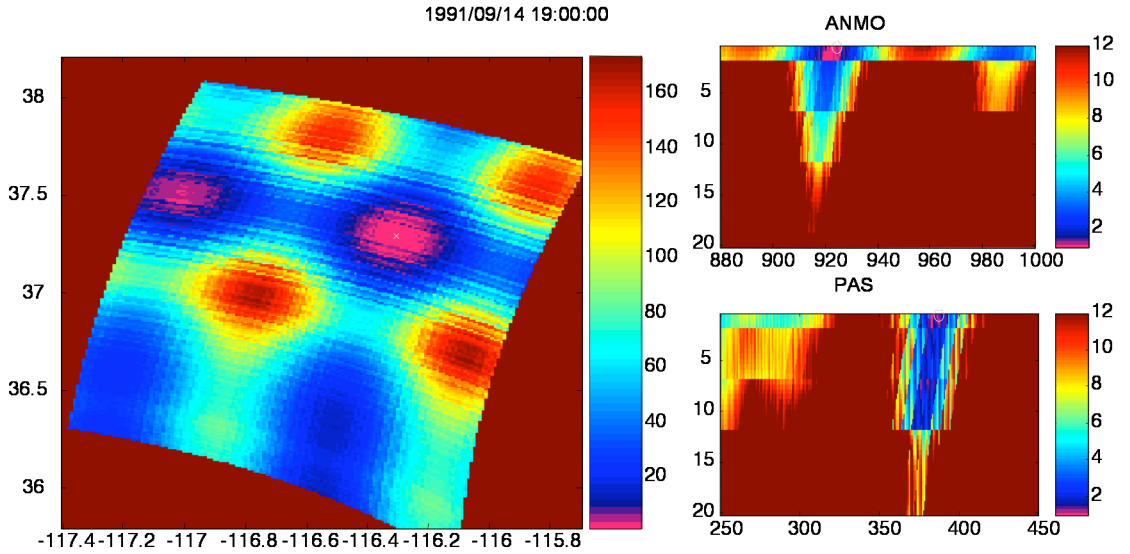
Figure 10: Range vs. Depth for ANMO and PAS (right) and a map view of the residual at the optimal depth (left) for the HOYA explosion. The Ground truth location is the white X.
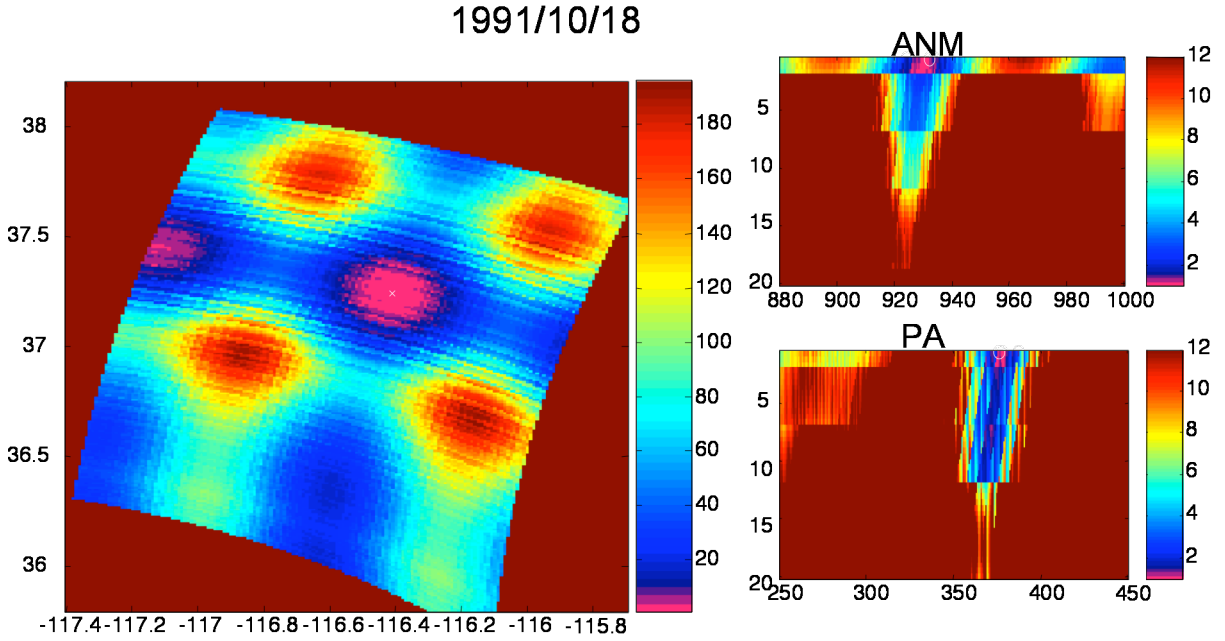
Figure 11: Range vs. Depth for ANMO and PAS (right) and a map view of the residual at the optimal depth (left) for the LUBBOCK explosion. The Ground truth location is the white X.

Figure 12: Range vs. Depth for ANMO and PAS (right) and a map view of the residual at the optimal depth (left) for the JUNCTION explosion. The Ground truth location is the white X.

16

Figure 13 Range vs. Depth for ANMO and PAS (right) and a map view of the residual at the optimal depth (left) for the Little Skull Mountain earthquake. The Ground truth location is the white X.

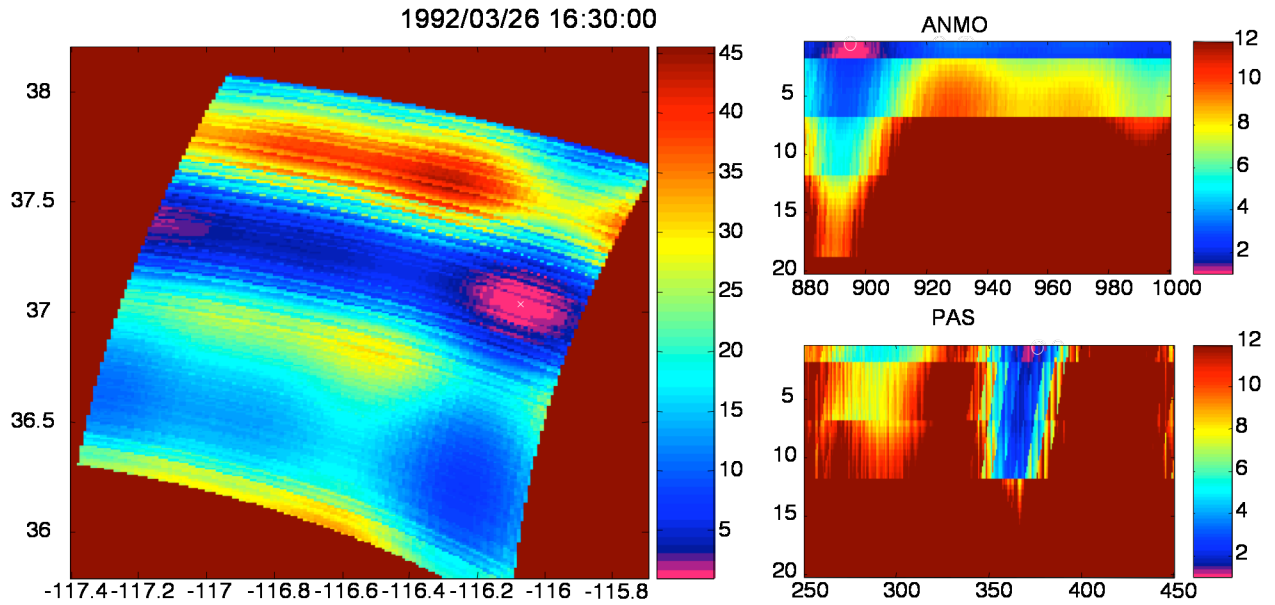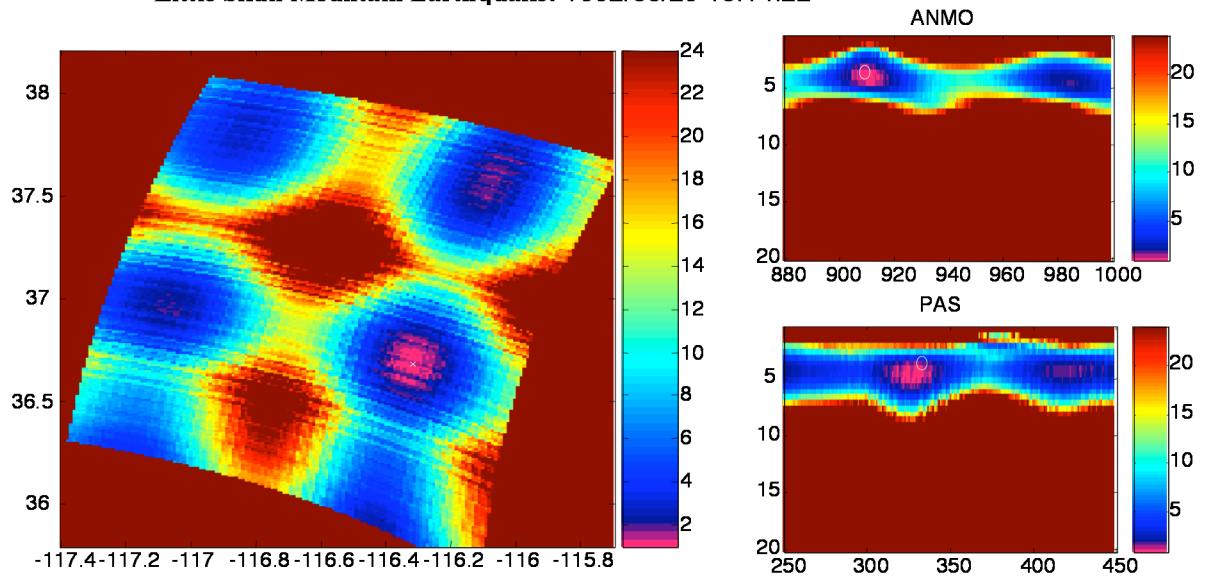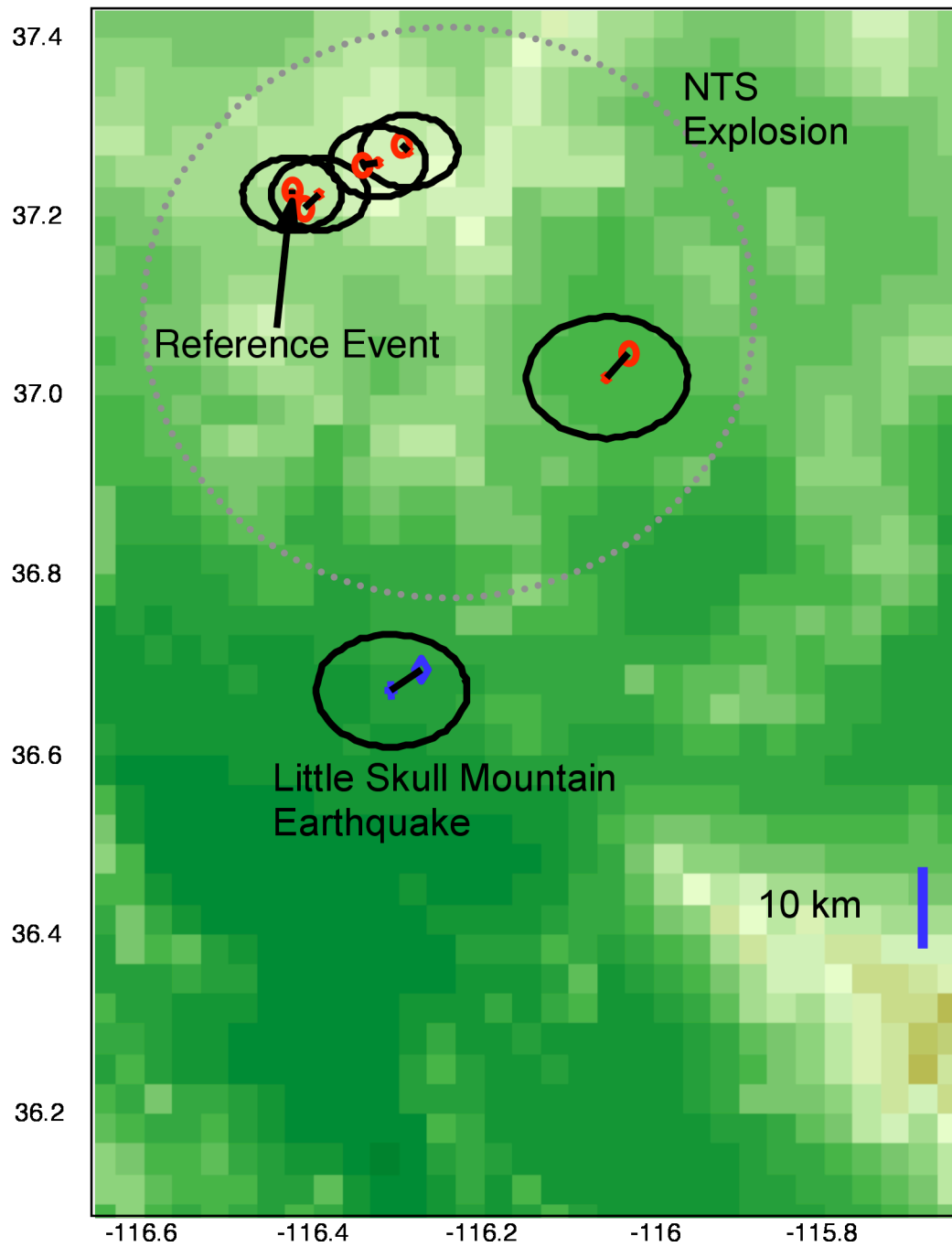**Figure 14: Map of the results for Nevada (x), the ground truth locations, (o), and the area of uncertainty (black ellipse). The blue line on the right is a 10 km scale.**
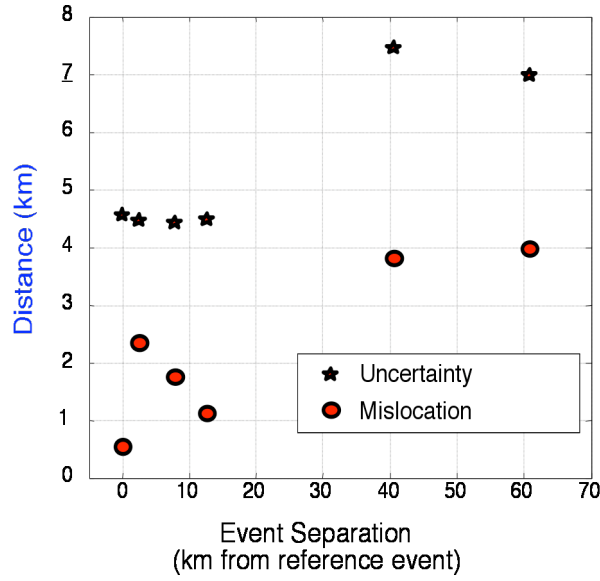
Figure 15: The error vs. ground truth, and the uncertainty ellipses compared with the offset for the new events and the reference event. Note that, in all cases, our locations were close to GT (always within 5 km), and consistently better than our estimates of uncertainty.

4.1.3.2. Range Dependent Example: Central California

As a second test case of the range-dependent semi-empirical synthetic locations, we attempted to relocate 55 events from Central California, shown in Figure 16 using the Mw=4.9 Parkfield, CA event of 2004/09/29 17:10:04 as a reference event. The purpose of this example is two fold: 1) demonstrate the approach over a larger data set, and 2) determine the maximum event separation for the approach. As the timing relative to the initial P-wave arrival is used to lag the surface waves, accuracy of the timing is critical. Unfortunately, the P-waves from the smaller events at the distant station, ELK, were emergent. To visualize this, Figure 17 shows the SNR of the P-wave plotted against the event magnitude and event-station range.

The primary impact of the low SNR is an increase in location error; Figure 18 shows the location error for each arrival when compared with the P-wave SNR. It is clear that, at high SNR's (>3), the location errors are small (<5km). However, at lower SNR's, the location error varies significantly. What is comforting, though, is that predicted uncertainty (from the RMS field) varies with the location error (Figure 20). In fact, of the 50+ measurements, the error is greater than the uncertainty in only one case. Since our confident level is 95%, it is reasonable to expect the errors to be greater than the uncertainty about 5% of the time.

19

After factoring data availability and data with poor SNR, accurate event locations were obtained for 22 of 55 processed events, as show in Figure 19 and Figure 21. Locations for many of the other events failed because of poor SNR for the P-wave arrival; some failed because the events were outside the range window of pre-computed Green's functions. The only factor in the ability to obtain quality locations is the SNR of the P-wave. Using other techniques, such as a correlation detector (for the P-wave) may allow for accurate locations for even smaller events than the magnitude 3.7 events in this data set. Equally important is that only two for the 22 events failed to reach GT5.
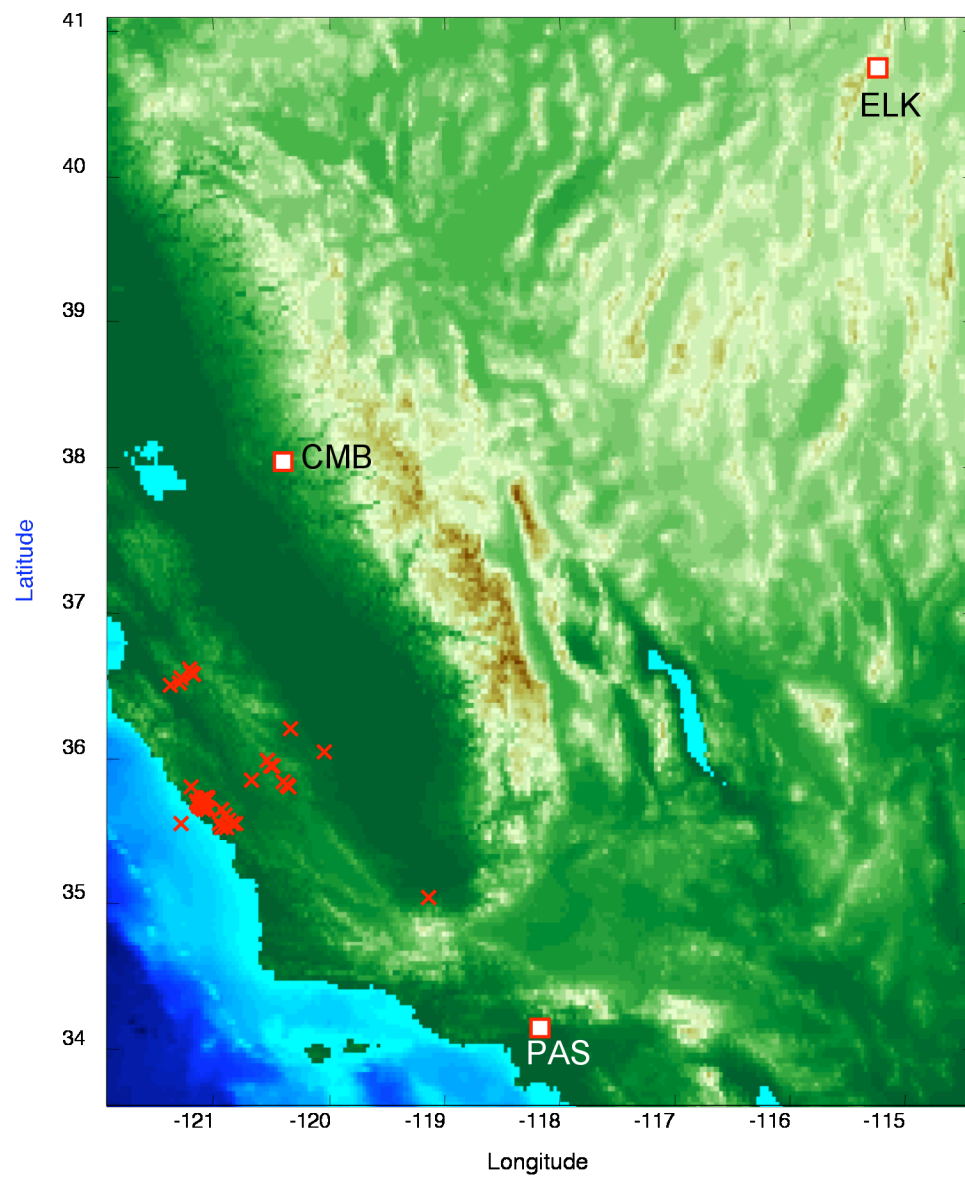
Figure 16: A map of the earthquakes we attempted to relocate and the stations used in the relocation.
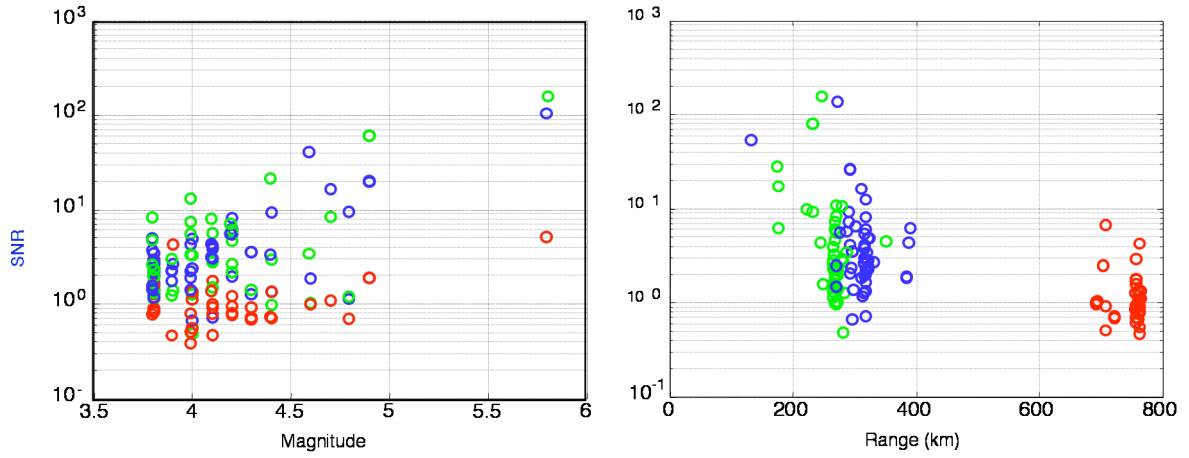
Figure 17: Signal to noise ration of the P-waves compared with event magnitude (left) and event-station range (right). The green circles are CMB, blue are PAS, and red are ELK.
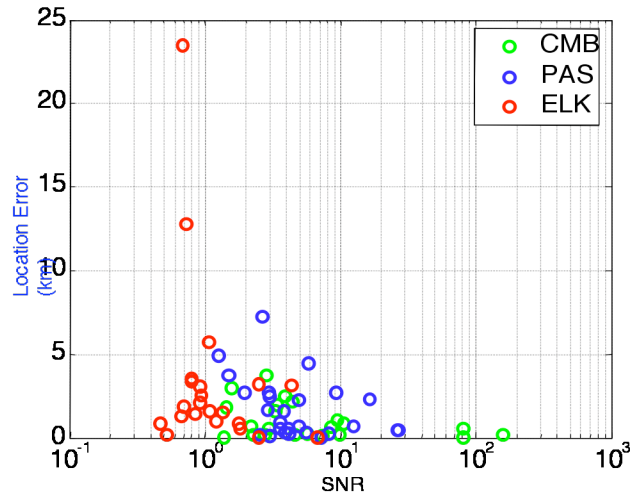
Figure 18: The location error in range for each arrival compared with the signal to noise ratio.
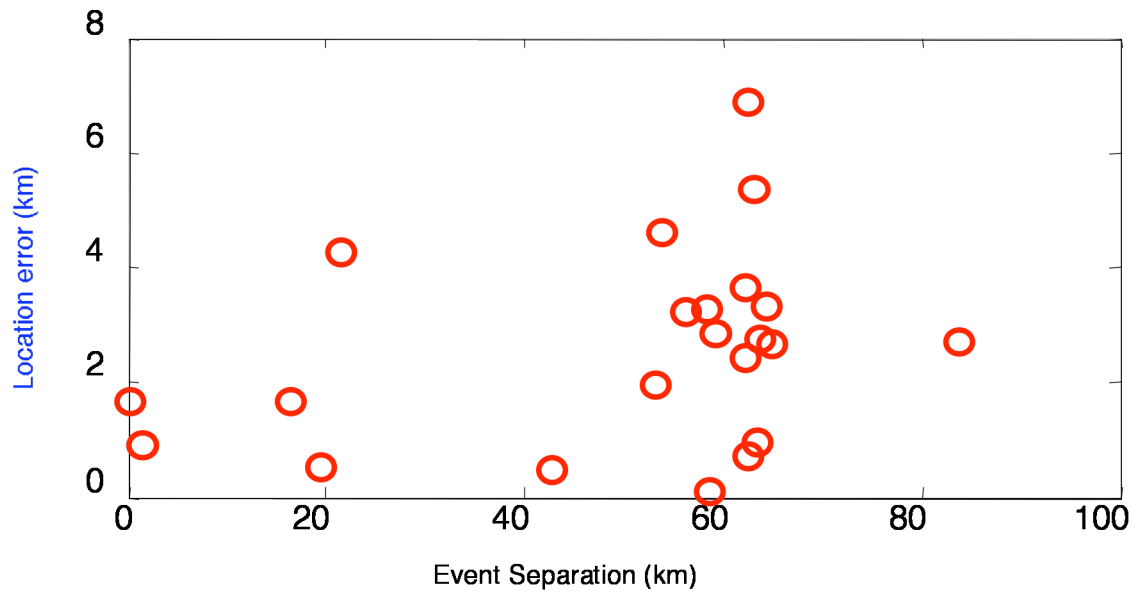
Figure 19: Mislocation distance between our determined location and the ground truth location compared with the separation between the reference and new events.
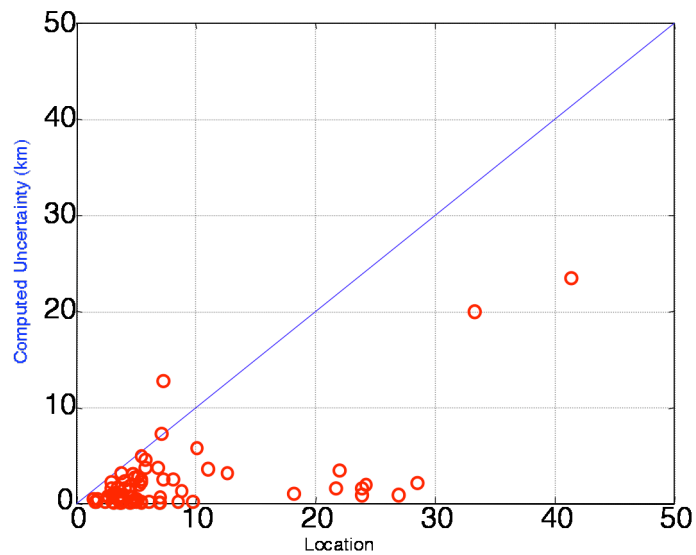
Figure 20: Computed uncertainty based on the RMS surface compared with the location error relative to ground truth. In most cases (all but 1), the computed uncertainty was greater than the error, which is consistent with a 95% confidence estimate
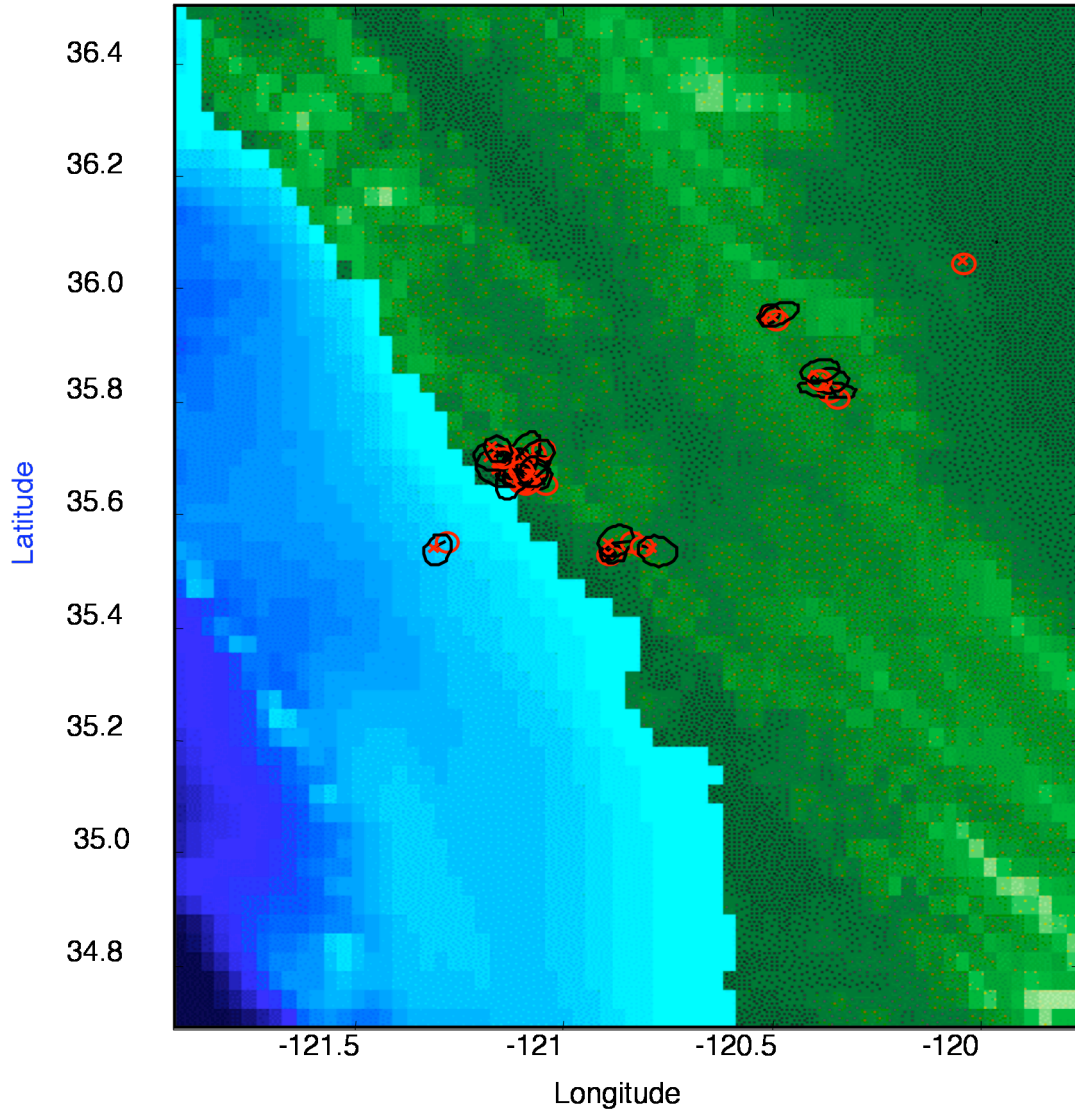
Figure 21: Map showing the ground-truth locations (red circles) and our locations and uncertainties (black).

4.1.3.3. Lop Nor, China

The range-depended location procedure was also applied to events in the vicinity of Lop Nor, China. The events studied include 5 Chinese Nuclear Explosions and one earthquake. The explosions are all GT1, according to Fisk (2002); the reported uncertainty in the earthquake location reported by the International Seismological Centre is better than GT5 (error ellipse is 3x4 km). Data is limited, as the Chinese stations did not transmit the data for the nuclear explosions. Therefore, the study is limited to stations outside of China. Two stations used: AAK and TL, as shown in Figure 22, and listed in

Table 2. With the five explosions, data was available at both stations for two of the events, and one (May 15, 1995) was the reference event. Single station range and depths were determined for the three events with only one stations, as well as an 'auto correlation' of the reference event to itself; the single station results are shown in Figures 23-27. These results indicate that we were able to accurately estimate the range (to within about 5 or 6 km uncertainty); the depth was resolved to be less than 20 km in all cases.

With the two events (one nuclear, one earthquake) where both stations were present, accurate (and precise) three-dimensional locations were obtained (latitude, longitude, and depth). The results, summarized in Table 3, and presented in more detail **Figure 28** and **Figure 29** for the explosion and earthquake, respectively.   In both cases, the locations were accurately (< 3 km from GT) obtained with precisions on the order of 5-7 km. Most significantly, though, is that the depths are clearly different, with the explosion being the shallower event. Examining the results in map view (Figure 30) demonstrates that the precision (size of the error ellipses) and accuracy of the locations obtained using the grid search; these locations were within a few kilometers of 'ground truth'. Another interesting result is the tradeoff between location and depth, as shown in Figure 31, which indicates a tradeoff between location (latitude specifically) and depth. This suggests that if the event location can be determined using conventional means, then that can be used to constrain the depth to within a few kilometers.
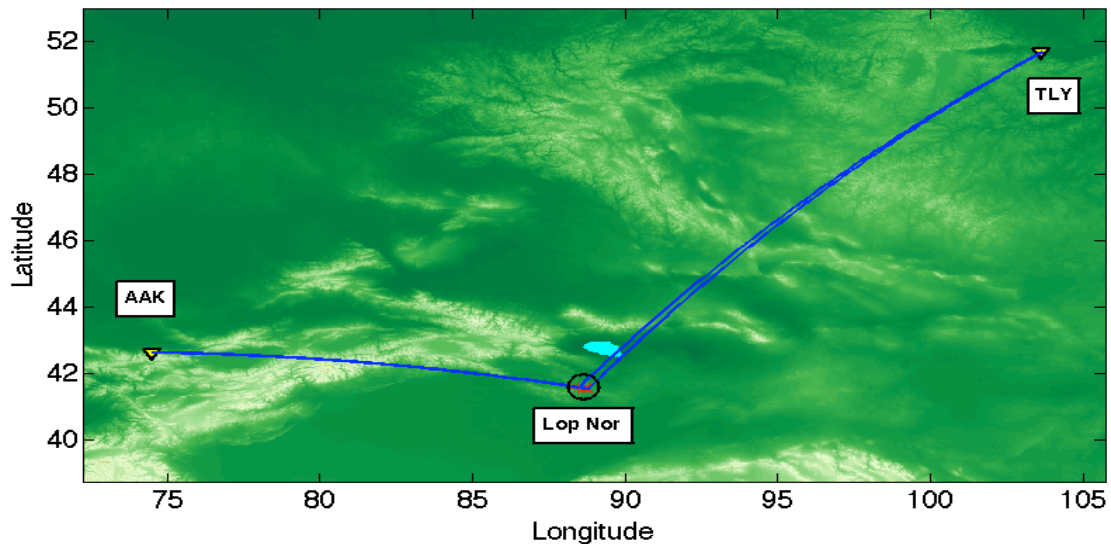
Figure 22: A map showing the location of the stations TLY and AAK relative to the Lop Nor test site. The event-station distances are about 1190 km to AAK and 1590 to TLY.
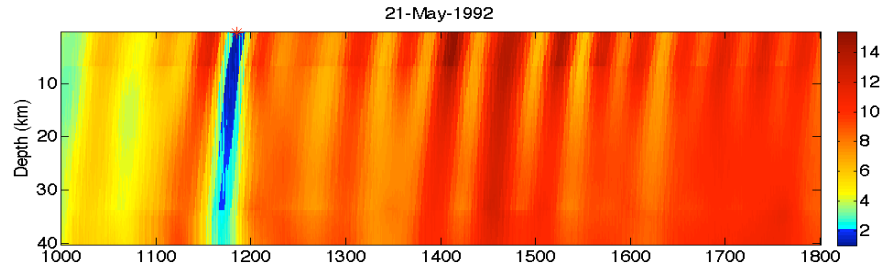
Figure 23: Range and depth at AAK for the May 21, 1992 Lop Nor test.
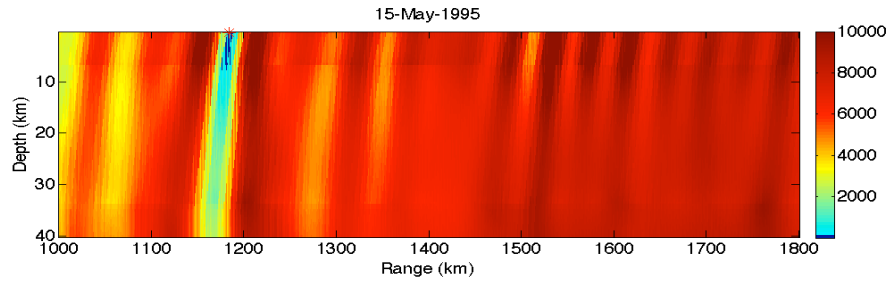
Figure 24: Range and depth at AAK for the May 15, 1995 Lop Nor test.
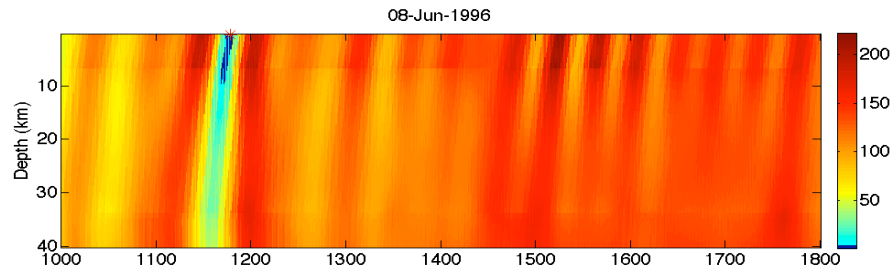
Figure 25:   Range and depth at AAK for the Jun 8, 1996 Lop Nor test.
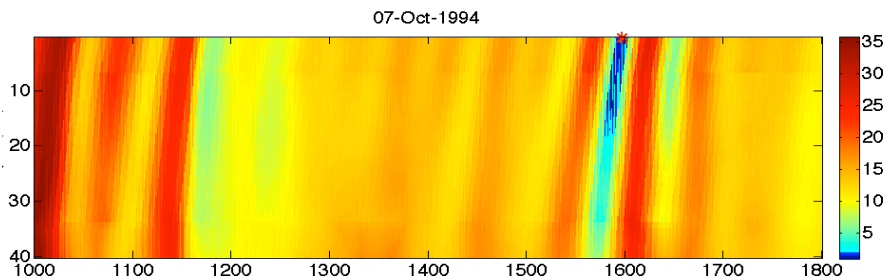
Figure 26: Range and depth at TLY for the Oct 7, 1994 Lop Nor test
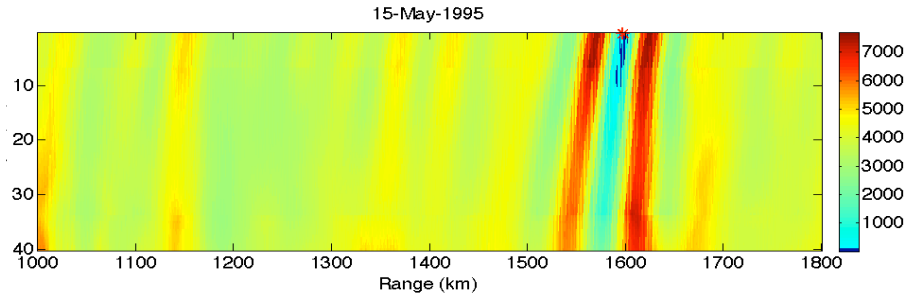
Figure 27: Range and depth at TLY for the Oct 7, 1994 Lop Nor test
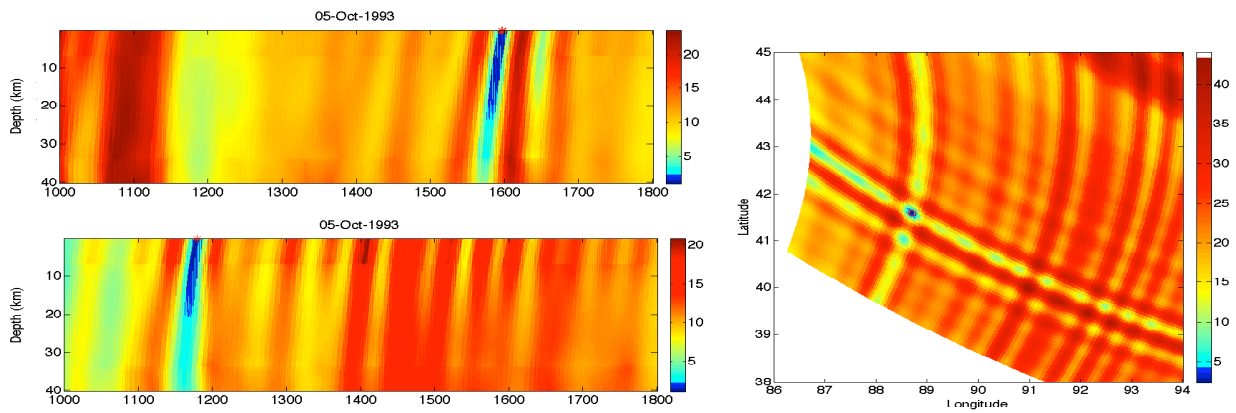
Figure 28: RMS as a function of range versus depth for the October 5, 1993 Lop Nor Nuclear Test. The results for station AAK are on the lower left, and TLY on the upper left. The red * represents the ground truth location. The maps on the right show the 2-D surface (RMS) at the 'optimal' source depth. Note that we did resolve to source to be shallow (near the surface).

27
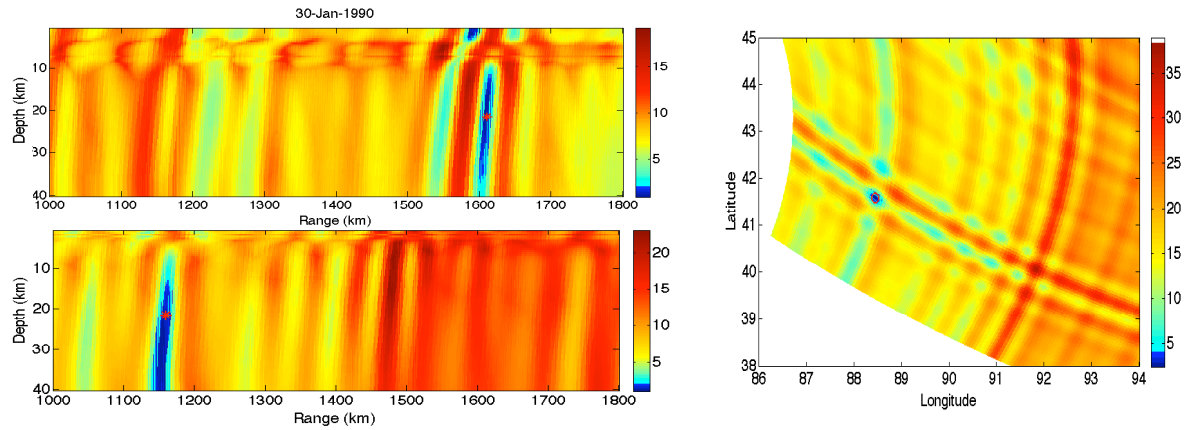
Figure 29: The RMS as a function of range versus depth for the earthquake that occurred on Jan 30, 1999 recorded at both stations. The results for station AAK are on the lower left, and TLY on the upper left. A red * represents the ground truth location. The maps on the right show the 2-D surface (RMS) at the 'optimal' source depth. Note that the optimal depth is resolved to be deeper than 15 km (not at the surface).
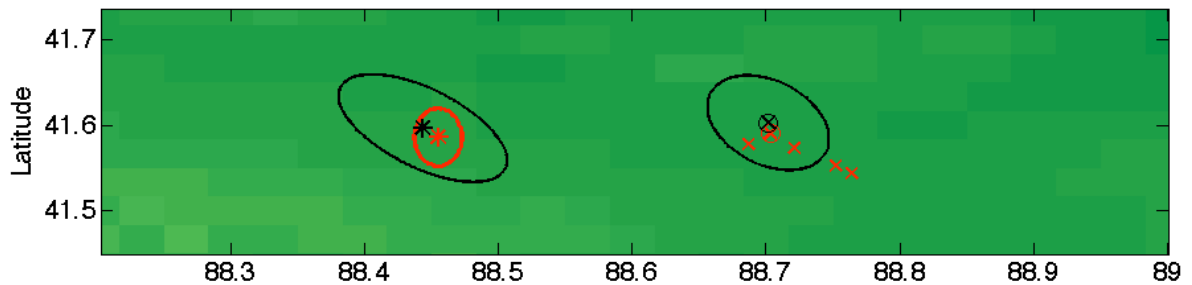
Figure 30: Our locations and uncertainties (black) compared with the ground truth results (red). While range vs. depth was computed for all events shown, only the reference event, one earthquake (*) and one explosion (x in circle) had data from both stations and could be located. Note the locations with two stations are within a few km of the 'ground truth' locations, even though our error ellipses (black) indicate that the errors are just outside of GT5.
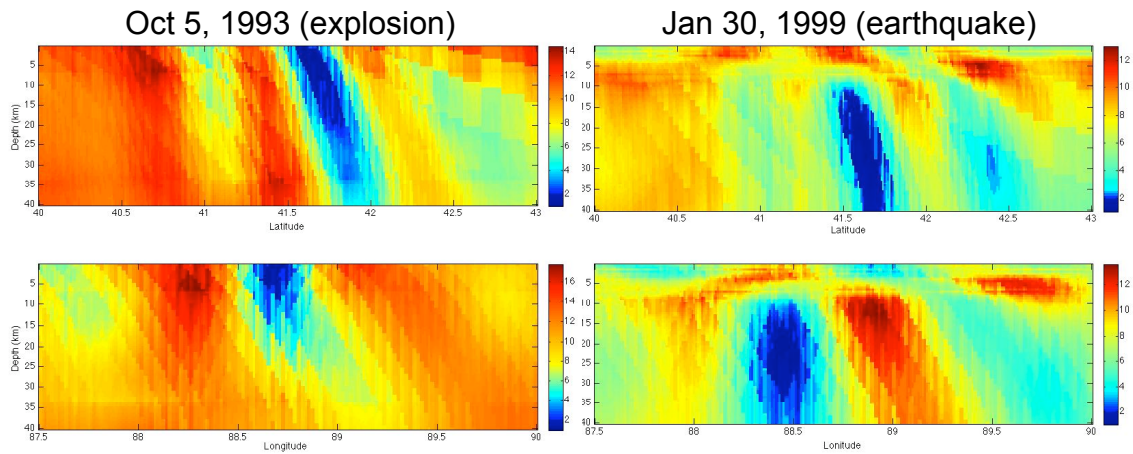
Figure 31: Tradeoffs between latitude (top figures) and longitudes (bottom figures) for the explosion (left) and the earthquake (right).

Table 2: Ground truth for the Lop Nor Events Analyzed. The first five events are Nuclear Explosions, with locations from Fisk (2001), the last event is an earthquake with the location from the ISC. The ground truth level (GT) for the earthquake is derived from the uncertainty in the ISC location.

EVID	Date	Lat	Lon	Dep	M_b	Stations	GT
1	1992/05/21 05:00:00	41.5437	88.7641	0	6.5	AAK	1
2	1993/10/05 02:00:00	41.5903	88.7031	0	5.8	AAK TLY	1
3	1994/10/07 03:26:00	41.5736	88.7209	0	5.9	TLY	1
4R	1995/05/15 04:06:00	41.5524	88.7524	0	6.0	AAK TLY	1
5	1996/06/08 02:56:00	41.5773	88.6869	0	5.8	AAK	1
6	1999/01/30 03:51:05	41.5860	88.4550	21	5.5	AAK TLY	~5

Table 3: Comparison of my grid search results with the ground truth results. My locations were assumed to be the center of the error ellipse derived from the RMS surface.

Evid	Best		My Loc		Error	My Error Ellipse			Depth
	Lat	Lon	Lat	Lon	(km)	Smaj	Smin	Az	(km)
2	41.5903	88.7031	41.59	88.70	0.26	5.95	4.45	-45	0
6	41.5860	88.4550	41.61	88.44	2.94	11.40	6.67	-55	23.5

4.2. Explosion Yield: Example from North Korea

We have used a slight modification of our semi-empirical synthetic seismogram method of Salzberg (2005; 2006) to estimate the yield of the North Korea Nuclear test of Oct 09, 2006. The data used for the processing is from the seismic station, MDJ, which is located in Northeastern china, as shown in **Figure 32**. The reference data, used for calibration, was from a wide-angle refraction experiment. The 1.2 – 1.5-ton shot was set off 190 km south of MDJ, or about ½ way between the North Korean test and MDJ. While the data for the Chinese reference event is noisy (Figure 33), at higher frequencies bands, (> 1 Hz), the signal is clear. The signal for the North Korea Nuclear test has significant signal at all frequencies (**Figure 34**).
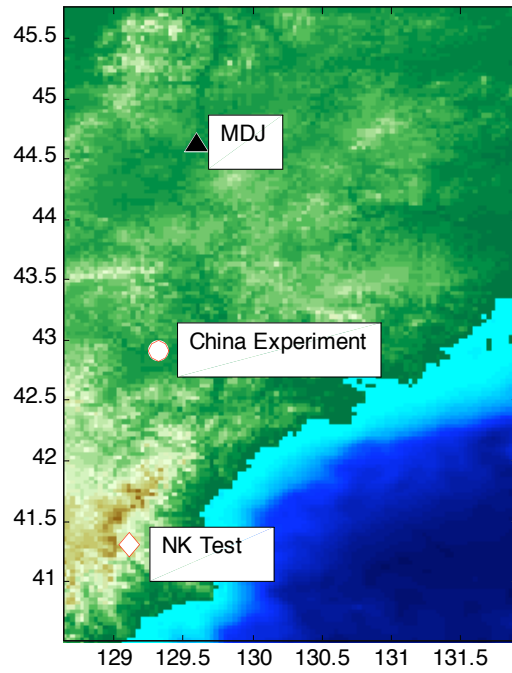
Figure 32: A map showing the relative locations of the North Korea test, the China experiment, and the seismic station MDJ.
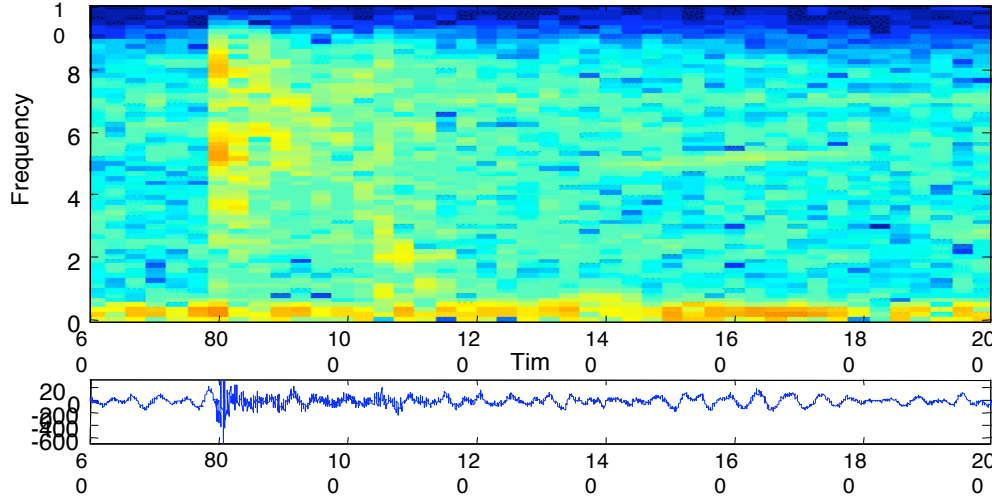
Figure 33: Unprocessed waveform and spectrogram for the waveform from a wide-angle refraction/reflection experiment in North Eastern China. (1.2-1.5 t). This event is used as a reference event.
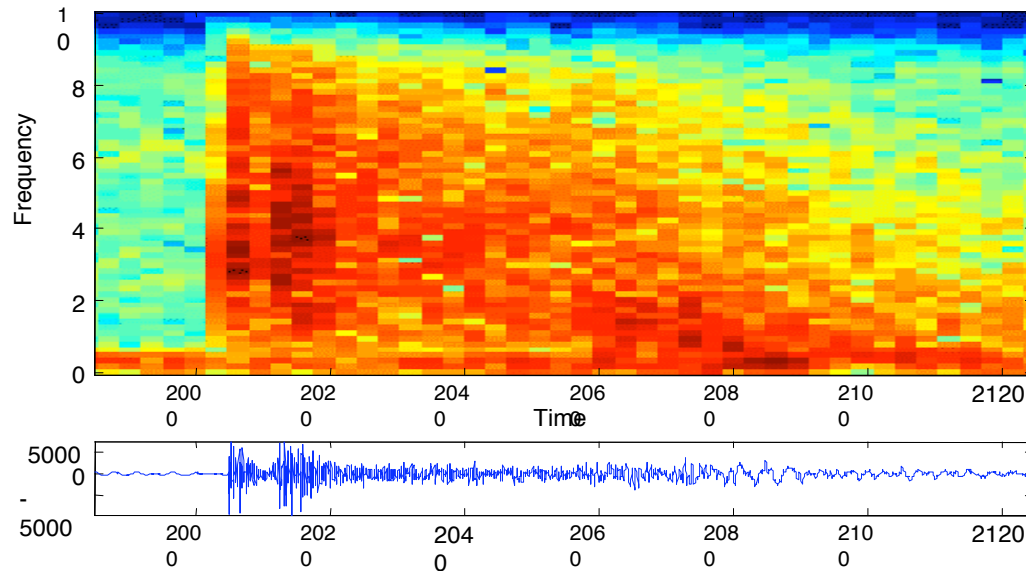
Figure 34: Unprocessed data for the North Korean Nuclear Test of Oct. 9, 2006.

As the separation between the reference event (Chinese experiment) and the North Korean explosion was significant (190 km, or ½ of the propagation distance, shown in **Figure 32**), and the frequency content of the data required high-frequency (>5 Hz) analysis, a coherent comparison was not feasible. Instead, the yield will be estimated by integrating (or summing) the energy envelopes. The data (High-pass filtered at 2 Hz) for the two events are shown in Figure 35. Conceptually, this approach can be viewed as

comparing the integrated energy envelops of the data (Figure 36) and synthetics (**Figure 37**). This is represented as:

$$Yield_{NK} = Yield_{CN} \cdot \frac{\sum O_{NK}}{\sum O_{CN}} \cdot \frac{\sum S_{CN}}{\sum S_{NK}}$$

Where O is the energy envelop for the observations of the two event, and S is the energy envelope of the synthetics for the two events, and the subscript NK refers to the North Korean test, and CN refers to the China experiment. The yield results are shown in Table 1, are obtained by multiplying the synthetic factor (3.7) by the observation difference (50) by the yield (1.2 – 1.5 tons), which gives 222 to 277 tons.
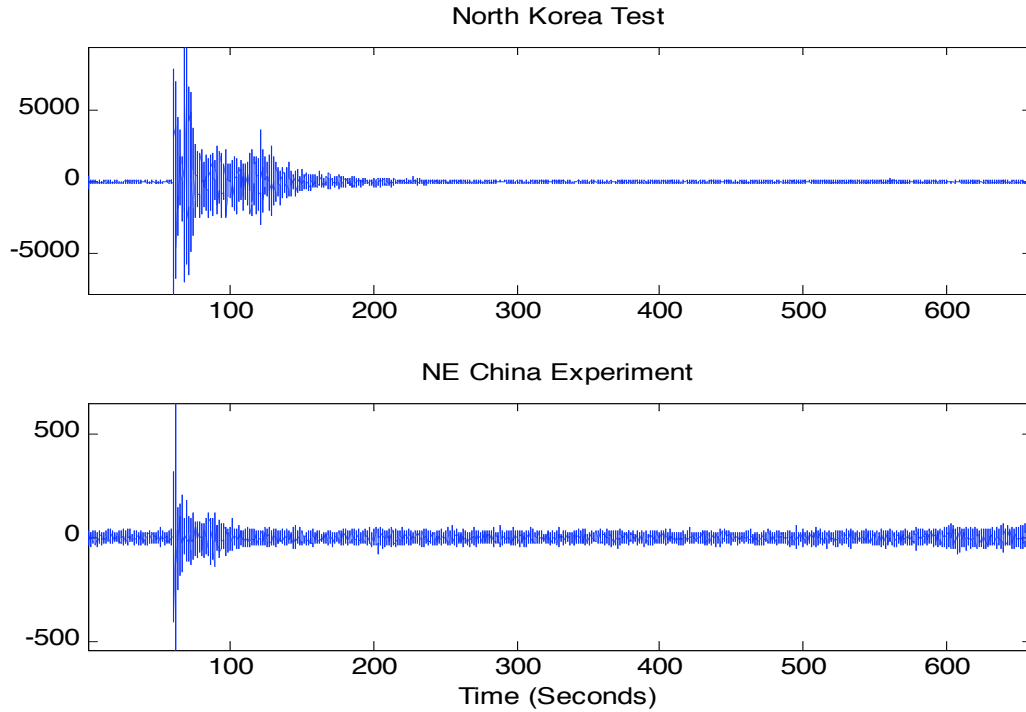
Figure 35: High-pass filtered waveforms at 2 Hz for the North Korean test and the Chinese experiment. The high pass filtering significantly enhances the signal-to-noise ratio.
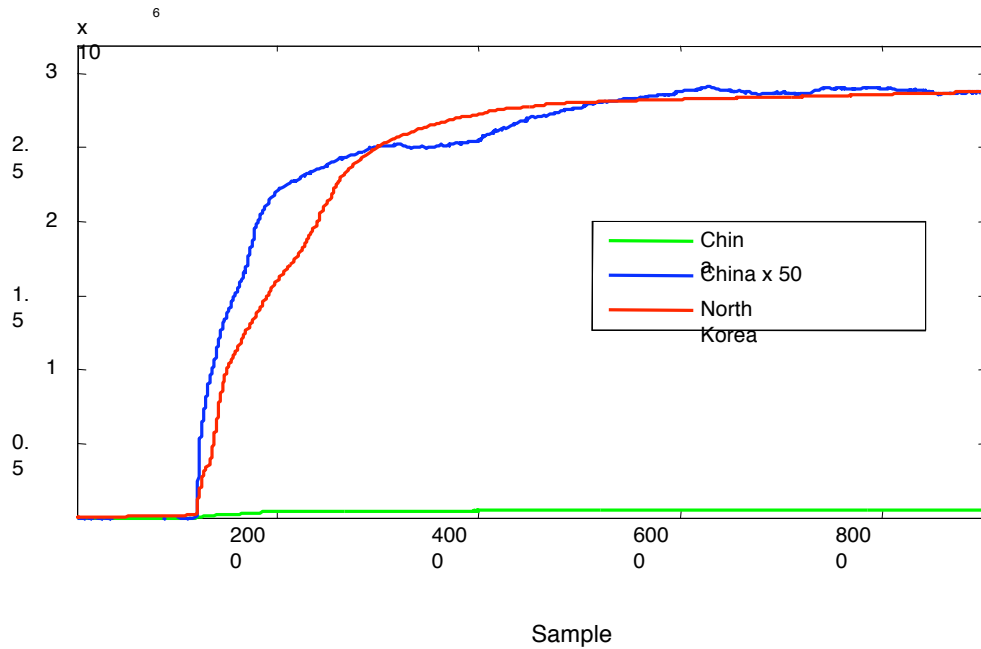
Figure 36 The integrated energy envelope for the waveform data from the North Korean test (red) and the China experiment (green), and the scaled China experiment (blue). A scaling factor of 50 was used.
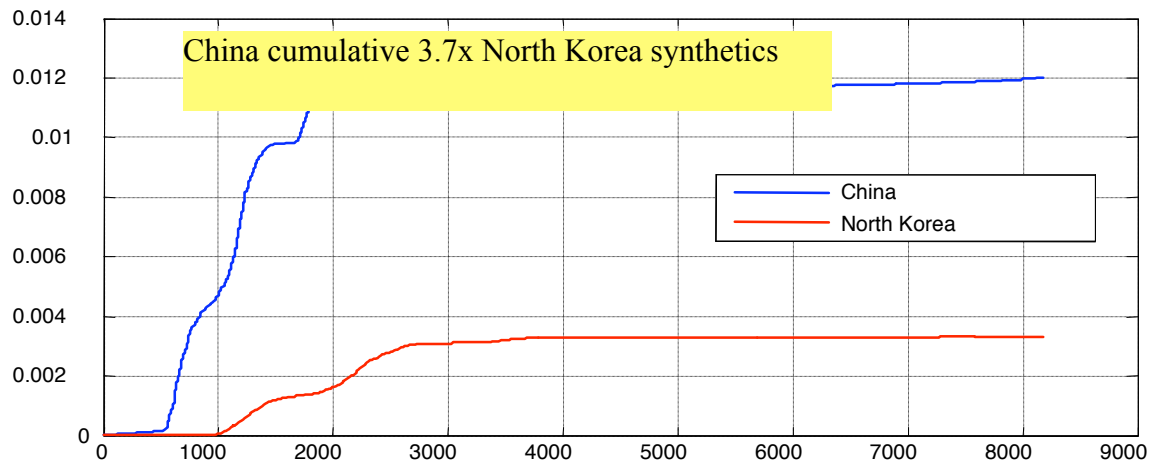
Figure 37 The integrated energy envelope for the synthetic waveforms from the North Korean test (red) the China experiment (blue). The difference of the curves results from the differing event-station distances (390 vs. 190 km).

Applying the empirical approach, and using synthetic seismograms that were computed using the IASPEI velocity model (Kennett and Engdahl, 1991) and Herrmann's (2002) wavenumber integration software for source depths of 0.5 km at ranges of 190.5m (China) and 369.5 km (North Korea). The empirical filtering was then used to transform the synthetic for China, and a scaling factor was determined by 1) windowing both the empirically filtered synthetic and the observed data using a window based on the signal to noise levels, and 2) summing the windowed envelopes.

Noise reduction
As an alternative using the high-pass filtering to improve the signal-to-noise ratio of the China test, we investigated using a Savitzky-Golay smoothing filter (Orfanidis, 1996) to characterize the noise. The noise is then subtracted from the observed waveform for further processing. The results, shown in Figure 38, indicate that this approach can be used to minimize the longer period noise.   The reduced long-period noise will allow for the processing at longer periods. These results are shown in **Figure 39**, and listed in Table 2.
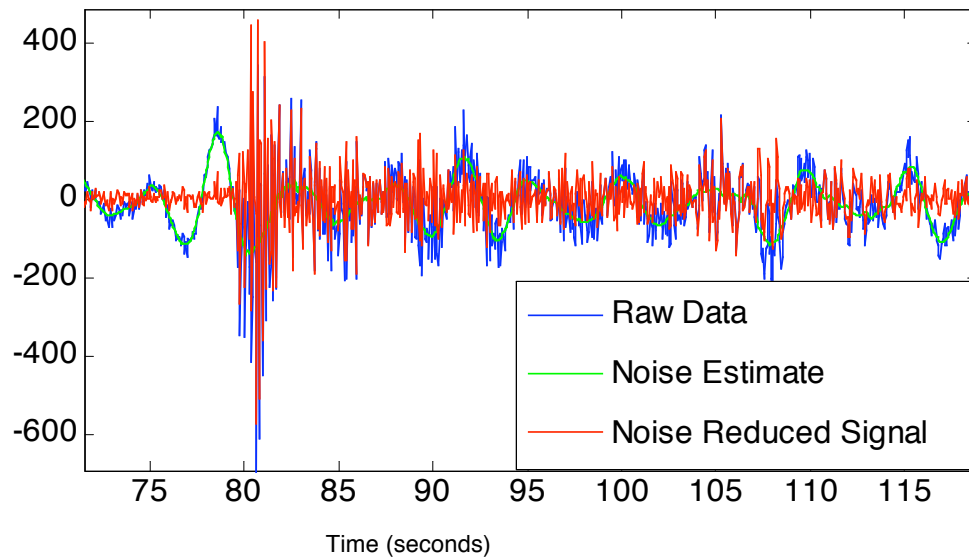
Figure 38: The raw data (blue), the noise estimate (green), and the residual (red) indicates that the Savitzky-Golay smoothing filter is successful in modeling the noise, allowing us to remove the noise from the data.
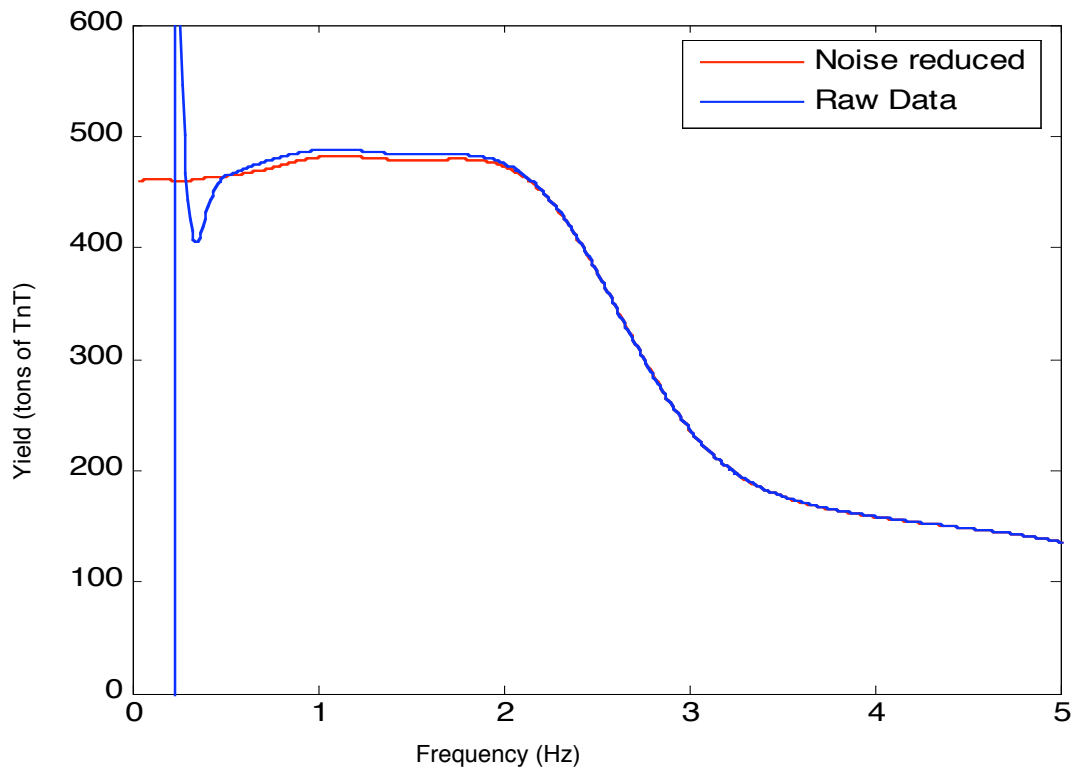
Figure 39: Yield estimates (in tons, vertical axis) based on the noise-corrected data as a function of high-pass filtered corner frequency. This shows, that at higher frequencies, the yield estimates are lower than a more broadband solution.

Table 4: Chemical Equivalent Yields with the different approaches.

Approach	Yield (1.2 t ref)	Yield (1.5 t ref)
Integrated Energy Envelopes	222	277
Semi-empirical, high-passed at 2 Hz, non-corrected.	213	267
Noise-corrected semi-empirical approach, broad band	372	465

5. CONCLUSIONS

In this and prior publications, we have demonstrated the utility of semi-empirical Green's functions for both source locations and source parameter extraction. In particular, we have showed that when a reference event and new event are nearly co-located, we can use a high frequency whole waveform approach to determine location to within ½ km. Once the two events are separated by more than a few kilometers, the differential propagation

velocities of the various seismic phases requires a phase-by-phase (wavenumber based) approach. With such an approach, we are able to locate events to within GT5 even when the events have different mechanisms and are separated by more than 50 km. Furthermore, we demonstrated the ability to use the approach in a variety of regions: Central California, NTS, and Lop Nor, China. We also demonstrated the ability to use the semi-empirical approach to determine source parameters (Yield) for the 2006 North Korean Explosion, even though the reference event was 180 km away.

REFERENCES

Aki, K. and P. Richards, <u>Quantitative Seismology, Second Edition</u>, University Science Books, Sausalito, CA, 2002.

Barker, J. S., <u>A Seismological Analysis of the May 1980 Mammoth Lakes, California Earthquakes</u>, Ph.D. Thesis, The Pennsylvania State University, State College, PA, 1984.

Burdick, L. J., and C. A. Langston, Modeling Crustal Structure Through the use of Converted Phases in Teleseismic Body-waveforms, *Bull. Seism. Soc. Am.*, **67**, 677-691, 1977.

Baggeroer, A. B., W. A. Kuperman, and Peter N. Mikhalesky, An Overview of Matched Field Methods in Ocean Acoustics, *IEEE Journal of Ocean Engineering*, **18**, 401-424, 1993.

Dziewonski, A. M., A. T. Chou, and J. H. Woodhouse, Determination of Earthquake Source Parameters from Waveform Data for Studies of Global and Regional Seismicity, *J. Geophys. Res.*, **86**, 2825-2852, 1981.

Fisk, M. Accurate locations of nuclear explosions at the Lop Nor Test Site using alignment of seismograms and IKONOS satellite imagery, *Bull. Seism. Soc. Am.* 92: 2911–2922., 2002.

Haskell, N. A., The Dispersion of Surface Waves in Multilayered Media, *Bull. Seism. Soc. Am.*, **43**, 17-34, 1953.

Helmberger, D. V., Generalized Ray Theory for Shear Dislocations*, Bull. Seism. Soc. Am.*, **64**, 45-64, 1974.

Herrmann, R. H., *Computer Programs in Seismology*, http://www.eas.slu.edu/People/RBHerrmann/CPS330.html, 2002.

International Seismological Centre, On-line Bulletin, http://www.isc.ac.uk, Internatl. Seis. Cent., Thatcham, United Kingdom, 2001.

Kanamori, H., and J. W. Given, Use of Long-Period Surface Waves for Rapid Determination of Earthquake Source Parameters 2. Preliminary Determination of Source Mechanisms of Large Earthquakes ($M_S > 6.5$) in 1980, *Phys. Earth and Planetary Int.*, **30**, 260-268, 1982.

Kennett, B. and Engdahl, E., Travel times for global earthquake location and phase identification, *Geophys. J. Int*. 105: 429–465, 1991.

Kikuchi, M. and H. Kanamori, Inversion of Complex Body Waves, *Bull. Seism. Soc. Am.*, **72**, 491-506, 1982.

Langston, C. A., Source Inversion of Seismic Waveforms: The Konya, India Earthquake of September 13, 1967, *Bull. Seism. Soc. Am,* **71,** 1-24, 1981.

Mendiguren, J. A., Inversion of Surface Wave Data in Source Mechanism Studies, *J. Geophs. Res.*, **78**, 889-894, 1977.

Menke, W. Geophysical Data Analysis: Discrete Inverse Theory, Academic Press, Orlando, FL, 1984.

Orfanidis, S. J., *Introduction to Signal Processing.* (Englewood Cliffs, NJ: Prentice-Hall), 1996.

Pulliam, J., C. Frohlich, and B. Phillips, Single Station Event Location: Epicentral Distance, Bearing, and Focal Depth, DOE Seismic Research Symposium, 2000.

Saito, R. Excitation of Free Oscillations and Surface Waves by a Point Source in a Certically Heterogeneous Earth, *J. Geophys. Res.*, **72**, 3689-3699, 1967.

Salzberg, David H., Simultaneous inversion of moderate earthquakes using body and surface waves: Methodology and applications to the study of the tectonics of Taiwan, Ph.D. thesis, State University of New York at Binghamton, Binghamton, NY., 1996.

Salzberg, David H., Karen E. Votaw, and Margaret E. Marshall, Seismic source locations and parameters for sparse networks by matching observed seismograms to semi-empirical synthetic seismograms, in *Proceedings of the 27th Seismic Research Review: Ground-Based Nuclear Explosion Monitoring Technologies,* LA-UR-05-6407, Vol. 1, pp. 454–462, 2005.

Salzberg, David H. and Margaret E. Marshall (2006). Seismic source locations and parameters for sparse networks by matching observed seismograms to semi-empirical synthetic seismograms, in *Proceedings of the 28th Seismic Research Review: Ground-Based Nuclear Explosion Monitoring Technologies*, LA-UR-06-5471, Vo 1, pp. 484–492, 2006.

Song, J., E.A. Hetland, F. T. Wu, X. Zhang, G. Liu, Z. Yang, P-Wave velocity structure under the Changbaishan volcanic region, NE China, from wide-angle reflection and refraction data. volcanic region, China. Tectonophysics, 433, 127-139, 2007.

Summerville, P. R. Graves and N. Collins, Ground Motions for Site Response Estimates – 1906 Earthquake, Technical Report to the Pacific Earthquake Engineering Research Center, 2000.

Velasco, A.A., C.J. Ammon, and T. Lay, Recent Large Earthquakes Near Cape Mendocino and in the Gorda Plate: Broadband Source Time Functions, Fault Orientations and Rupture Complexities, *J. Geophys. Research*, **99**, 711-728, 1994.

Wu, F. T., Prediction of String Ground Motion Using Small Earthquakes, *Proc. Of the 2nd International Conference on Microzonation,* National Science Foundation, San Francisco, CA, 701-704, 1978.

Yang, X., R. North and C. Romney, *CMR Nuclear Explosion Database (Revision 3)*, CMR Technical Report CMR-00/16, 2000.